Zuzana Parusniková
David Merritt *Editors*

Karl Popper's Science and Philosophy

Springer

Karl Popper's Science and Philosophy

Zuzana Parusniková · David Merritt
Editors

Karl Popper's Science and Philosophy

Springer

Editors
Zuzana Parusniková
Institute of Philosophy
Czech Academy of Sciences
Prague, Czech Republic

David Merritt
Rochester Institute of Technology
Rochester, NY, USA

ISBN 978-3-030-67035-1 ISBN 978-3-030-67036-8 (eBook)
https://doi.org/10.1007/978-3-030-67036-8

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2021
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Of all philosophers of the twentieth century, Karl Popper stands out as the one who did most to build bridges between the diverse academic disciplines.

His first major work, *Logik der Forschung* (1934), concerns scientific method. Popper's ideas were formed in the intellectual climate dominated by the logical positivism of the *Wiener Kreis*; despite a great diversity in academic interests, the members of the Vienna Circle wanted to reaffirm the scientific ethos of the Enlightenment ideal. Excited by the revolutionary ideas of Einstein (whom they engaged in both scientific and philosophical discussions), they believed that philosophy must play an active role in this new era by drawing as close to science as possible. Although Popper shared these general ideals, he strictly rejected all the main pillars of the positivist philosophy of science: inductivist logic of discovery, the verifiability principle and the concern with meaning. In single-handed opposition to this influential philosophical movement, Popper offered new solutions: a hypothetico-deductive view of science, based on falsifiability as the demarcation criterion and a denial of the claim that scientific theories could be verified. It is fair to say that the radicalism of Popper's proposals caused an upheaval among philosophers of science, especially after the publication of his work in English in 1959.

With the advent of World War II, Popper applied his revolutionary methodological ideas to political philosophy. He became famous for his theory of the open society, in which he criticized authoritarian and totalitarian social systems based on the doctrine of historicism, that is, historical inevitability. The future is open, said Popper, and since we all are fallible so are our social and political systems. Holistic experiments, a willingness to sacrifice one's life for a higher good, must be avoided and replaced by a more modest piecemeal social engineering, in which mistakes can be corrected and society reformed without bloodshed. The same is true, he argued, for political regimes: Popular replacement of governments is the keystone of democracy, and democracy is—despite its many imperfections—the best form of government known so far.

Later, Popper focused on wider problems of the growth of knowledge. Rational discussion, he suggested, depends on a readiness to listen to critical arguments and should not aim to demonstrate truth. Scientific theories are guesswork, but by constantly subjecting theories to testing, science can progress. His methodological principle of criticism is thus the core of a dynamic but challenging epistemology, requiring an adventurous spirit and a willingness to make risky conjectures. Falsification—Popper's "negative methodology"—takes on a positive role that of uncovering new problems through the elimination of failed hypotheses. Popper shifted the focus of methodology from proving to undermining, from establishing to critical activity itself. In a broader philosophical sense, he proposed an antifoundationalist model of rationality that views all knowledge as conjectural, hypothetical and provisional.

Not surprisingly, Popper is one of the few philosophers of science who inspired scientists (especially the Nobel Prize winners Peter Medawar, Jacques Monod and John Eccles, in addition to the biologist Donald Campbell, the biochemist Günter Wächtershäuser and the mathematician Hermann Bondi), and he won recognition by the scientific establishment (he was elected a Fellow of the Royal Society in 1976). It was Popper's emphasis on scientific research as an adventure, in which scientists constantly and fearlessly attack received opinions in the search for the truth and for new and interesting problems, that was so much admired.

This is not a list of Popper's contributions; let us nevertheless mention his herculean success in presenting an axiomatic system for probability that provides a genuine generalization of (propositional) deductive logic, his success in developing the theory of logic as a theory of deduction, his defense of realism in quantum mechanics, his study of the body–mind problem and his involvement in discussions of evolutionary biology. His methodology and epistemology have been widely and vividly discussed, but his impact on scientific research and his contributions to it have received less attention. The aim of this book is thus to illustrate, and evaluate, the impact, both substantive and methodological, that Popper has had in the natural and mathematical sciences. An attempt is made to pinpoint the connections between these contributions and his central philosophical concerns. The topics selected are quantum mechanics, evolutionary biology, cosmology, mathematical logic, statistics and cognitive science. The approach is multidisciplinary, opening a dialogue across scientific disciplines and between scientists and philosophers.[1]

It is always fascinating to watch the moments of rupture when philosophy acquires a completely new impetus and challenges the established ways of perceiving the world. Karl Popper overturned the traditional values ascribed to reason and revolutionized the field of philosophy of science. Inevitably, his views provoked debates and disagreements. Our own goal here is not to glorify Popper but to invite the study of his best ideas and develop critical perspectives through the evaluation of his ideas and his work.

Prague, Czechia — Zuzana Parusniková
Rochester, NY, USA — David Merritt

---

[1] The impulse to investigate Popper's influence on science emerged at the Popper Symposium held at the CLMPST 2019 in Prague and supported by the Karl Popper Charitable Trust. Many thanks to Joseph Agassi who contributed with valuable comments and advice to this volume.

Contents

Biology

Cognitive Science

Contributors

Joseph Agassi Tel Aviv University, Tel Aviv-Yafo, Israel;
York University, Toronto, Canada

Peter Århem Department of Neuroscience and Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

Nimrod Bar-Am Communication Department, Sapir College, Shaar Hanegev, Israel

David Binder Department of Computer Science, University of Tübingen, Tübingen, Germany

Constantin C. Brîncuş Institute for Research in the Humanities, The Research Institute of the University of Bucharest, Bucharest, Romania

Manjari Chakrabarty Visva Bharati (Central University), Santiniketan, India

Flavio Del Santo Faculty of Physics, University of Vienna, Vienna, Austria;
Institute for Quantum Optics, Quantum Information, Austrian Academy of Science, Vienna, Austria

Olival Freire Jr. Instituto de Física, Universidade Federal da Bahia, Salvador, Brazil

Nir Fresco Ben-Gurion University of the Negev, Beer Sheva, Israel

Helge Kragh Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

Anastasiia Lazutkina LUT University, Lappeenranta, Finland

Youngjo Lee Department of Statistics, Seoul National University, Seoul, South Korea

Philip Madgwick Milner Centre for Evolution Department of Biology & Biochemistry, University of Bath, Bath, UK

Olga Markič Department of Philosophy, Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia

David Merritt Rochester, NY, USA

Hans-Joachim Niemann Poxdorf, Germany

Denis Noble Department of Physiology Anatomy & Genetics, University of Oxford, Oxford, UK

Raymond Noble Institute for Women's Health, University College London, London, UK

Zuzana Parusniková Institute of Philosophy, Czech Academy of Sciences, Prague, Czechia

Yudi Pawitan Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Thomas Piecha Department of Computer Science, University of Tübingen, Tübingen, Germany

Michael Ruse Florida State University, Tallahassee, FL, USA

Yanhua Shih Department of Physics, University of Maryland, Baltimore, MD, USA

Karl Popper: His Philosophy and Science

Zuzana Parusniková

1 Knowledge and Criticism

Karl Popper is one of the few philosophers of science who is well-known to scientists and respected by them. Apart from the direct influence of his views on science (especially in the fields of physics, cosmology, logic, biology and the philosophy of mind) it is his methodology that most appeals to scientists. It is based on the "traditional" values endorsed by science: the confidence in reason and progress, realism and the conception of truth as correspondence with facts. Popper stands against certain fashionable trends by which the development of science can be viewed as a procession of discontinuous paradigms, as a collage of local discourses, or as a series of language games with their own internal and relative standards of truth.

But Popper earned the highest accolades from scientists for his emphasis on criticism as the essence of progress in science. His methodological principle of criticism is the core of a dynamic epistemology, requiring an adventurous spirit and expressing the high-risk nature of the cognitive process. Via the falsification of proposed hypotheses, the positive role of erring is exploited to discover new problems, leading to the growth of knowledge. As Popper puts it,

> … empirical science may be defined by means of its methodological rules. … First a supreme rule is laid down which serves as a kind of norm for deciding upon remaining rules, and which is thus a rule of a higher type. It is a rule which says that the other rules of scientific procedure must be designed in such a way that they do not protect any statement in science against falsification. (Popper 1965, 54)

While in his early writings Popper is more concerned about science and the criterion of demarcation between science and metaphysics, later he considers criticism in a broader philosophical context. He appreciates its key role in developing creative imagination in all areas of life and sees it as a means to discover new ideas and new

---

Z. Parusniková (✉)
Institute of Philosophy, Czech Academy of Sciences, Prague, Czechia
e-mail: parusnikova@flu.cas.cz

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_1

horizons of knowledge, as the ability of breaking through dominating frameworks of the epistemological, cultural and ideological discourses.

> It seems to me that what is essential to "creative" or "inventive" thinking is a combination of intense interest in some problem … with highly critical thinking; with readiness to attack even those presuppositions which for less critical thought determine the limits of the range from which trials (conjectures) are selected; with an imaginative freedom that allows us to see so far unsuspected sources of error; possible prejudices in need of critical examination. (Popper 1982a, 48)

In contrast, the desire for justification is a defensive attitude that prevents us from daring guesses and thus inhibits our willingness to risk. Instead, low-risk but unchallenging conjectures with low empirical content are favoured in the hope that they will not be refuted by tests. Such an attitude, though, suppresses intellectual conflicts and their creative potential. As Popper argues, justification and content are inversely related and therefore the aim of achieving the highest possible probability of our statements leads to a timid approach, generating uninteresting conjectures; "the probability of a statement (or a set of statements) is always the greater the less the statement says" (Popper 1985, 58). Simply put: the less we say the less we err. Justificationist strategies, which aim at verifying or confirming a theory, at establishing its highest possible certainty, reliability and credibility, stifle human curiosity and sabotage the growth of knowledge. However, if we take falsifiability is a methodological goal we are not afraid of hazarding bold (improbable) guesses in the hope of arriving at new, deeper problems. The critical approach is a booster while justification is an inhibitor of creative energy.

1.1 The Dangers of Dogmatism

Popper's requirement of permanent criticism became one of the key issues in the Popper-Kuhn controversy; this debate took place at the International Colloquium in the Philosophy of Science in London in 1965 and, together with Imre Lakatos' launching the conception of scientific research programmes, defined some of the subsequent lines of criticism of Popper. Kuhn defended his view on the development of science as a succession of discontinuous paradigms, presented in his book *The Structure of Scientific Revolutions* (1963); a change of paradigms marks a revolution in science, generated by the long-term inability of ordinary research to deal with anomalies, and results in a fundamental restructuring of scientific concepts, methods and standards. Once the new paradigm gets established this revolutionary phase is followed by a relatively long period of "normal science", in which scientists build up the potential of the paradigm and are busy with puzzle-solving rather than with serious criticism. In Kuhn's words, "their work produced not simply new information but a more precise paradigm, obtained by the elimination of ambiguities that the original from which they worked had retained. In many sciences, most normal work is of this sort…" (1970, 34).

For Popper, this tedious, uninspiring (allegedly) normal science is an antithesis of what science should be; against Kuhn he claims that "science is essentially critical; it consists of bold conjectures, controlled by criticism, and it may, therefore, be described as revolutionary … 'the normal' scientist, as Kuhn describes him, is a person one ought to be sorry for" (Popper 1982b, 55, 52). Kuhn, however, insists that "observation cannot … force you to falsifying conclusion, and you would occasionally be the loser if it could do so " (Kuhn 1982, 18). Popper argues that Kuhn's "normality" leads to dogmatic thinking that lures us to hold on to beliefs using all kinds of justificationist trickery, such as ignoring counter-evidence and marginalizing the warning signs when our expectations, based on a certain hypothesis, are not fulfilled, when crucial tests uncover inconsistencies in the proposed hypothesis. For Popper, "a dogmatic attitude is clearly related to the tendency to *verify* our laws and schemata by seeking to … confirm them, even to the point of neglecting refutations" (Popper 1985, 50). Whereas Kuhn considers "grand" revolutions in science, such as represented, for instance, by Copernicus , Newton and Einstein , Popper views all scientific activity as consisting in permanent (though often less spectacular) "micro" revolutionary cognitive acts.

Popper equates the rational foundations of science with criticism. Unlike Popper, Kuhn deals with actual scientific practice, including historical, sociological and political aspects—for instance, how scientific community demands commitment from scientists to certain research goals within the ruling paradigm, their loyalty to a "common intuition" that resembles a tribal initiation act; especially young scientists are supposed to show their allegiance to the "board of elders" in their institutions, selling their free spirit for grants and jobs. As Watkins observes, "Kuhn sees the scientific community on the analogy with a religious community and sees science as the scientist's religion " (Watkins 1982, 33) and, similarly, Fuller argues that "Kuhnian normal science … combines the qualities of the Mafia, a royal dynasty and a religious order " (Fuller 2003, 46). It may be debated to what extent this description reflects the typical practice in scientific communities and institutions or just picks up on some extreme cases. Popper himself, though, is not interested in sociological issues but in the internal logic of the growth of knowledge—the driving force of which, as emphasized several times, is criticism. Normal science by its very nature—defined by justification demands in testing and by the demands for conformity to authorities—feeds on dogmatism. For Popper, that is exactly the reason why Kuhn's *Realpolitik* of science must not be sanctioned as a methodological norm and rather serves as a deterrent example of the corruption of rationality; normal science, Popper says, is "a danger to science and, indeed, to our civilization" (Popper 1982b, 53)

This danger is, according to Popper, so acute since dogmatic tendencies are inborn in our nature; we are instinctively predisposed to expect regularities in Nature and want them confirmed. We find confirmation reassuring, giving us a (false) feeling of security. These expectations, Popper argues against Hume , are not formed inductively by repeated experience of certain regularities, but are biologically a priori. Thus, *dogmatism is a biological force that is anchored in our genes.* Not only, then, is there a dogmatic streak in our nature, but it is a primal instinctive force. As Popper says, “this dogmatic attitude, which makes us stick to our first impressions, is indicative of a strong belief; while a critical attitude … is indicative of a weaker belief” (Popper 1985, 49). The ability of criticism is a privilege of the human species that emerged only at a higher stage of our evolution; namely at the stage when humans developed the argumentative function of language, encouraging abstract thinking, and above all writing; the written word enabled us to achieve a higher level of critical culture.

1.2 Criticism Needs a Boost

In his *Objective Knowledge*, Popper introduces his ontological conception of three worlds; World 1 contains physical entities, World 2 represents the world of subjective cognitive acts and processes, and World 3 is the world of intelligibles, objective knowledge, written and stored. Once problems and conjectures become linguistically formulated they can be communicated, shared and *criticized*. When conjectures are written down the debates and criticisms can be carried out even more efficiently than if they are just uttered. However, this level of criticism is not available to all: as Popper remarks, “dogmatic thinking, an uncontrolled wish to impose regularities, a manifest pleasure in rites and in repetition as such, are characteristic of primitives and children” (Popper 1985, 49); the term “primitives”, as I understand it, refers to tribal formations, such as the “cold cultures” described by Lévi-Strauss . A higher degree of intellectual refinement leads to less dogmatism and to the development of the imaginative, reflexive and critical form of intelligence. Conversely, a lack of intellectual maturity implies stronger dogmatism and the prevalence of biological instincts in our attitudes to the world.

Popper’s conception of evolutionary epistemology and its link to evolutionary biology are not the issues of this chapter but are discussed in Part III of this book. I want to emphasize just one point crucial for the understanding the role of criticism for Popper. The existence of World 3 gives humans an unprecedented evolutionary advantage. Since the evolution of knowledge takes place in the world of abstract entities we can delegate the survival struggle to our ideas. In case of animals, fatal errors result in death or, worse, the extinction of the species—the elimination of errors occurs in World 1, the world of physical entities. But we, humans, have another option—we can let ideas die in our stead. True, this option is not always favoured and we still keep killing each other because of ideological disagreements, but the possibility of the choice itself is unique. And the right choice implies a positive attitude to erring—*since we do not fear for life we can actively pursue criticism and learn from our mistakes*. To use Popper’s famous example, “the difference between the amoeba and Einstein is that although both make use of the method of trial and error-elimination, the amoeba dislikes erring while Einstein is intrigued by it: he consciously searches for his errors in the hope of learning by their discovery and elimination” (Popper 1979, 70).

Popper exposes a fundamental clash between our dogmatic nature and critical reason, especially as the dogmatic inclinations affect our spontaneous cognitive attitudes. *But it is criticism that is the weak link and needs a boost*; it is a fragile and relatively recently acquired ability constantly confronted with our robust dogmatic predisposition. Such a boost can be provided via the methodological imperative that defines criticism as the sole means of assessing conjectures.[1] And although Popper's philosophy of science does not deal much with sociology and politics of scientific practice, the Popperian message, consisting in an appeal for the institutional endorsement of criticism on all levels of research and for a positive evaluation of scientists who discover errors in their own hypotheses, can be applied to this area, too.[2] Not silly errors, of course, but such that arise from an expert knowledge of the field, from serious attempts to solve pressing problems and that inspire new lines of investigation.

2 From Demarcation to a New Concept of Reason

Popper´s new concept of reason unfolded from his dispute with the logical positivism of the Vienna Circle. The Circle was the center of intellectual life in between the Wars and the cradle of analytical philosophy; chaired by Moritz Schlick it gathered philosophers, mathematicians, physicists, economists and social theorists and represented a progressive movement that stood in opposition to various speculative and conservative tendencies in philosophy and politics, and that reaffirmed the scientific ethos of the Enlightenment ideal. Logical positivists lived and worked in the time of Einstein's discoveries and were engaged in both scientific and philosophical communication with him; they believed that philosophy must play an active role in this new era by drawing as close to science as possible.

Inspired by Wittgenstein, the philosophers of the Vienna Circle developed the method of logical analysis of language by which only those statements which can be empirically verified are scientific: philosophical language must be purged of all metaphysical elements. The immediate experience is verified by its sense-giveness and expressed in the simplest, atomic statements (protocol sentences). More complex

---

[1] Popper, however, sometimes compromises this imperative by claiming that a certain amount of dogmatism in the initial phase of assessing a new hypothesis is necessary—a theory should show its mettle before it is submitted to the destructive effect of criticism; "a critical attitude needs for its raw material, as it were, theories or beliefs which are held more or less dogmatically" (Popper 1985, 50). I see this claim as dangerous. Any declaration of the legitimacy or even rationality of dogmatism—however limited it may be—tames criticism. However, in critical rationalism a methodological norm must not tame criticism or it consequently loses its regulative appeal. In the whole context of Popper´s writing I regard it as a lapse, in which Popper overlooks his own arguments concerning the primordial power of dogmatism (Parusniková 2017, 27–39).

[2] The need of cultivating an open-minded atmosphere in scientific community is discussed for instance by Agassi and Jarvie (1987): scientific workshop provides the ideal environment in which "critical debate about open questions, employing tentative answers which are accepted either as grounds for action or as agenda for further debate or both" (44).

statements are built according to the logical rules of syntax, as suggested by the titles of Carnap's two most famous books, *Der logische Aufbau der Welt* (1928) and *Logische Syntax der Sprache* (1934), and these statements must be inversely reducible to the atomic statements from which they are inferred.[3] This was, according to the positivists, the criterion of demarcation between science and pseudoscience. Historically, logical positivists drew on the kind of empiricism and the inductive method developed by Hume and Mach : all ideas ("concepts", for Mach) must be grounded in experience or they are fictional[4]; in Hume´s words, "all our ideas or more feeble perceptions are copies of our impressions or more lively ones"; and if not, "they are not founded on reasoning, or any process of the understanding" (Hume 2005 [1748], 19, 15).

Popper also felt the need for defining a criterion of demarcation and he, too, endorsed empiricism. But the impossibility of obtaining an empirical proof of the validity of inductive inferences was for him a reason to deny the validity of induction on logical grounds. This is of course an old epistemological problem: the ancient sceptics pointed out the circularity of reasoning involved in justification; Hume argued that inductive reasoning presupposes the validity the principle of induction or, in Hume's words, the Uniformity Principle. This could be done either by proof (by demonstration) or by probability (reasoning from experience). The former is impossible since proof requires certainty while the Uniformity Principle is contingent; "'tis possible that the course of nature may change, since we can conceive such a change " (Hume 1981 [1739], 651). Neither can experience provide proof because it too is predicated upon the uniformity of nature. As Hume remarks, "[A]ll probable arguments are built on the supposition, that there is this conformity betwixt the future and the past … this conformity is a *matter of fact*, and … will admit no proof but from experience. But our experience in the past can be a proof of nothing in the future" (ibid., 651–652). This verdict is also acknowledged by Russell; "[The] principle itself cannot, of course, without circularity, be inferred from observed uniformities, since it is required to justify any such inference" (Russell 1946, 699).[5]

Logical positivists were, of course, aware of this problem and came up with various probabilistic solutions with the hope of bypassing the problem of the legitimacy of inductive inferences, using the method of confirmation instead of verification. Popper, again following Hume , rejected this solution, "[F]or if a certain degree of probability is to be assigned to statements based on the inductive inference, then this will have to be justified by invoking a new principle of induction" and so on ad infinitum (Popper

---

[3] References to logical positivism take into account neither the differences between individual philosophers nor the changes in their views over time. I present only the constitutive features defining their positions.

[4] In full, Hume´s copy principle goes as follows: "All our simple ideas in their first appearance are deriv'd from simple impressions, which are correspondent to them, and which they exactly represent " (Hume 1981 [1739], 4).

[5] For instance, in seeking to prove the principle of induction, Russell tries to avoid this deadlock as follows: "we shall nevertheless hold some knowledge is a priori in the sense that experience which makes us think of it does not suffice to prove it, but merely so directs our attention that we see its truth without requiring any proof from experience" (1912, 116).

1965, 30). We are in a closed circle and the only two choices left are infinite regress or the dogmatic acceptance of the principle of induction.

As is well-known, Popper claimed to have solved this problem by replacing the criterion of verifiability by falsifiability. Empirical testing consists in formulating singular existential statements (basic statements) that state the existence of at least one observable event forbidden by the hypothesis, and then carrying out experiments designed to falsify the hypothesis (as discussed in the famous Mill/Hempel/Popper examples of black ravens and white swans). Popper thus defined a new criterion of demarcation: "But I shall certainly admit a system as empirical or scientific only if it is capable of being *tested* by experience. These considerations suggest that not the *verifiability* but the *falsifiability* of a system is to taken as a criterion of demarcation" (Popper 1965, 40).

Popper then proceeded to reject induction not only as an incorrect logical inference but as an inaccurate description of knowledge acquisition. Knowledge does not start with pure observations, he claimed, the instruction "observe" is absurd; when he asked his students: "take a pencil, carefully observe and write down what you have observed", the first question was "what should we observe?" (Popper 1985, 46). Observation is, according to Popper, always selective and theory/problem laden. Popper thus turns the positivist model upside down; he declares that induction is a myth in both the logical and the psychological sense, adding, with his typical audacity, that it is he himself who must admit responsibility for the death of logical positivism (Popper 1982a, 88).

Like almost everything that Popper proposed, his model of the growth knowledge as fueled by the falsification of high-risk (improbable) hypotheses has also given rise to heated debates. Of those I choose two themes that can best show Popper´s shift from the narrower logic-oriented discourse to a new philosophical vision of rationality. Firstly, it is the status of basic statements and secondly, his concept of corroboration. Popper´s treatment of these two issues open the space for *defining reason in negative terms, as an agent of destruction of knowledge claims*, and to *viewing all knowledge as conjectural, hypothetical and provisional.*

2.1 Basic Statements

One of the most frequent objections to Popper´s conception of basic statement as potential falsifiers of a theory is as follows: in order to falsify a hypothesis, we would have to prove the truth of (i.e. justify) the basic statement that is supposed to falsify the hypothesis. However, this cannot be done without employing induction since even basic statements contain universals—and, as the argument goes, if they are not verifiable no decisive falsification is possible; thus Popper´s criterion of demarcation does not work.

This objection misinterprets Popper; he never claimed that basic statements can be justified: "we do not attempt to *justify* basic statements [by perceptual experience]. Experiences can *motivate a decision*, and hence an acceptance or rejection of a statement, but a basic statement cannot be justified by them" (Popper 1965, 105). We accept—not prove—a basic statement is true through a decision or by agreement that are motivated by the empirical evidence available at the time; still, though, basic statements remain open to further examination if or when required. Thus for Popper even these most elementary observational statements, serving as the basis for the acceptance or rejection of a theory, do not have a foundationalist nature. Likewise, of course, the refutation of a theory remains conjectural: "All knowledge remains fallible, conjectural. There is no justification, including, of course, no final justification of a refutation. Nevertheless, we learn by refutations, i.e. by the elimination of errors, by feedback" (Popper 1988, xxxv).

Popper´s conception of falsification thus marks a huge shift away from the traditional emphasis on proofs in science and offers a different vision consisting in hypothetical reasoning and encouraging the approach of permanent dissent. Yet even though he rejects all kinds of justification, which displays itself in the efforts to ground knowledge in some ultimate authority, in some unquestionable certainty from which we can then proceed to build science, Popper does not resign on truth as the regulative goal of science; he just divorces truth from justification. Therefore, his dissident approach is not a sceptical game, such as we see in the ancient sceptics; in contrast he says that "rational discussion must not be practised, however, as a mere game to while away our time. It cannot exist without … the search for objective truth…" (Popper 1988, 157).

2.2 Corroboration

Similar misinterpretation occurs in the case of Popper´s concept of corroboration. Defined in deductive terms, corroboration means that *the theory has been tested and has not yet been falsified*: "we say that a theory is ´corroborated´ so long as it stands up to these [genuinely critical] tests" (Popper 1965, 266). Corroboration is a minimalist, ascetic concept—it merely states that basic statements have not so far contradicted the tested theory, and further attempts to give it a try are encouraged. It does not entail any empirical support of the theory or any suggestion concerning its future performance; and when Popper talks about degrees of corroboration he does not mean any degrees of justification but only the severity of tests that the theory has undergone. Theories with high degrees of testability are rich in content and thus have high potential to provoke critical discussion. It is once again obvious that Popper's main concern is the *process* not *proof*; unlike positivist confirmation, corroboration does not establish any firm foundations of knowledge in the sense of probability, reliability or justified belief. However, corroboration motivates our decision to accept the theory as provisionally true or as the best approximation to the truth.[6]

---

[6] I leave out Popper´s controversial concept of the degrees of corroboration as linked to the degrees of verisimilitude; this is discussed in Part II of this book.

Many Popper critics raise the objection that corroboration gives us too little; we need more and any ‘more’ would make corroboration a weaker form of confirmation. If based purely on *modus tollens* corroboration does not have any informative or predictive value, they say; and if it does, then it has to employ induction (e.g. Salmon 1968; O’Hear 1980; Newton-Smith 1981). In other words, corroboration has to add something to the theory, some support that the theory did not have before. How else could we justify our preference between theories, our decision to act on the selected theory and our belief that it will succeed? John Watkins (1984, chap. 8) summarizes these concerns and asks: why do corroborations matter? Other questions then follow: is the best corroborated (most severely tested) theory always the best theory? What reason, if any, is there for an agent to act on best corroborated theories? To answer these and other similar questions some Popperians, such as Lakatos and Musgrave , have argued that it is necessary to incorporate some inductive principle into critical rationalism, or science would be left in chaos—any preference for certain theories or decisions in applied science and technology would be a matter of uncontrolled guesswork bordering on irrationalism; therefore, some philosophers of science suggest separating the theoretical and the pragmatic problems of induction. All these positions assume that corroboration should deliver more than a mere lack of refutation: if the tested theory is not falsified it is fortified.

Such attempts to squeeze some positive assurance from corroboration fail to appreciate the novelty of Popper´s imperative of falsification based on the principle that reason is incompatible with justification. In order to understand Popper´s radical shift away from the traditional theory of knowledge one should quit demanding that it delivers the traditional wares. In this respect I support David Miller´s dismissive approach regarding the attempts to mitigate the negative definition of corroboration; Miller (1994) argues that questions of the sort mentioned above beg for a justificationist answer and try to draw Popper into the very game the rules of which he rejects. For Popper, the method of science “relies on expulsion procedures, rather than entrance examinations”, says Miller, and “the expulsion procedures are the sole means that Popper allows for the control of scientific knowledge” (ibid., 6–7). Corroboration, therefore, does not add anything to the theory; the theory enters the testing arena, in which it is being mercilessly attacked, *already with the claim to be true*; this claim is either rejected or (provisionally) retained.

Knowledge thus remains in the realm of guesswork, yet our guesses are not blind but controlled by criticism. Popper´s concept of corroboration thus stands in total opposition to positivist confirmation and is free from any justificationist elements. This poses a significant challenge for science, requiring a change of perspective. Retaining a theory (or a solution to a problem) that has survived criticism is a rational decision, yet lacks any guarantees.

3 The Popperian Challenge

Although scientists appreciate Popper´s appeal to criticism as the engine which drives scientific progress they may be wary of the radical implications of the falsificationist imperative; at least many philosophers of science seem to be. As discussed above, confirming evidence is of no value for Popper. The Popperian challenge entails that the attempts to verify or confirm a theory—apart from being irrational—bring no benefits to the growth of knowledge. This stands against the deeply ingrained conviction held by traditional epistemology, requiring that a theory be supported by evidence. Obviously, it is not easy to part with this tradition and many philosophers of science raise criticism against Popper on this point, claiming that some evidential support is essential in the appraisal of theories. In order to have the best of both worlds, they say, this support should be taken into account only if it comes from a genuine attempt to refute the theory.

3.1 Anti-Justificationist Extremism

Such strategies, though, try to smuggle a justificationist element into critical rationalism and as a result bring Popper´s concept of corroboration (as “not yet falsified”) too close to the positivist confirmation. By definition, an imperative does not allow any compromise. Any mitigation of Popper’s critical imperative would result in a banal appeal to be critical. In other words, in a philosophical rehash—nobody would dispute that it is useful to scrutinize hypotheses for errors. Popper’s radical falsificationism would become half-hearted, uncontroversial advice: let us be critical and suspicious until proved otherwise—until the theory is not vindicated. The justificationist flavour achieves the reverse—it rids Popper of all taste and transforms his challenging position into something trivial.

Similarly, it may be hard to come to terms with Popper´s view that even refutations remain conjectural—there is no conclusive proof or disproof of any statement; even potential falsifiers of a theory cannot be verified. True, we have to stop the testing process at some point but that is, for Popper, a matter of a decision, as risky as any other can be—and it can be reversed if necessary. The game remains open on all levels since science does not rest upon the bedrock. Popper compares science to a building erected on piles. “The piles”, he says, “are driven down from above into the swamp, but not down to any natural base; and when we cease our attempts to drive our piles into a deeper layer, it is not because we have reached firm ground” (Popper 1965, 111). The picture of science balancing on poles in a swamp may not seem too alluring; I therefore prefer Watkins´ metaphor of theories “floating in the ocean of uncertainty ” (Watkins 1984, 354).

The Popperian challenge implies that *uncertainty marks the terrain of rational inquiry and forms our ultimate noetic predicament*. Yet Popper encourages us to embrace this predicament instead of fearing it—or viewing it as defective in science.

Living in uncertainty, yet not giving up the search for truth, is thrilling. It makes us shift our attention from proving to seeking, from establishing to problem solving, exploiting the creative potential of our thought. Popper traces this ideal back to Xenophanes of Colophon, and takes the following passage as his philosophical motto[7]:

> Through seeking we may learn and know things better.
> But as for certain truth, no man has known it,
> Nor shall he know it, neither of the gods.
> Nor yet of all the things of which I speak.
> For even if by chance he were to utter.
> The final truth, he would himself not know it:
> For all is but a woven web of guesses.

3.2 *Scientific Practice*

Another set of objections to critical rationalism targets its suitability for scientific practice: based on various case-studies from the history of science the Popperian methodology is deemed unrealistic. This position was first posited by Thomas Kuhn and Imre Lakatos , followed by some of Popper´s other colleagues from the LSE plus other collaborators. These critics argue that scientific theories cannot be tested in the same way as statements in logic; they are clustered together, forming interconnected theoretical systems or, in Lakatos´ words, research programmes. These resist the falsifying impact of a single piece of empirical counter-evidence, mainly for two reasons. The first is utilitarian—the resistance to refutation is beneficial to science if a research programme has rich heuristic power (including increasing empirical content, successful prediction and the production of novel knowledge). Worrall takes this position to the extreme (and approaches close to Kuhn´s concept of normal science) when he claims that "the history of science shows that scientific progress is best made (perhaps only made) *not* by holding every assumption equally open to criticism …" (Worrall 1996, 97) . The second line of criticism targets the very possibility of falsification: if a prediction deduced from a theory proves wrong we cannot exactly determine what is falsified. Drawing on the Duhem-Quine thesis, this argument shows that the core theory is linked to other theories and to a number of auxiliary assumptions, all of which can be blamed for failing the test—in sum, and using Worrall´s expression again, science is messy, too messy for falsification to work as a methodological norm.

These debates have undoubtedly opened up new fields of problems concerning the application of falsificationism in the complex developments and practice of science. In this context, they show both the vitality of critical rationalism and its possible shortcomings. However, they also raise the question of whether Popper is not, perhaps,

---

[7] Fragment 34 from Sextus Empiricus; the first line (sometimes added) is fragment 18 from Stobaeus.

expected to deliver too much—to provide an exact account on the history of science and to give a precise recipe for how to proceed in every appraisal of complex units of theories. Well-designed experiments may help, but, of course, errors may occur anywhere along the line when deciding which part of the theory or which auxiliary assumption is falsified. The implementation of the critical discourse also depends upon the willingness of the scientists to specify crucial tests that would refute the theory.[8] In the *Introduction* to his *Realism and the Aim of Science* Popper gives a list of twenty examples from the history of science when refutations led to revolutionary theoretical reconstructions; he also admits that sometimes it may take time before the refutation is accepted. Yet in the same breath he adds that his methodology is not intended to be a historical theory but a *philosophical normative proposal*. This is very important. Popper's radical falsificationism should not be interpreted as a descriptive history or sociology of science. His critical imperative is strictly normative. Popper postulates that criticism is the core of rationality and therefore it ought to be the means by which our conjectures are evaluated.

3.3 New Concept of Reason

As I said, Popper may be expected to deliver too much in terms of his account of the history of science, yet quite how much he delivers in philosophical terms can tend to go unappreciated. *The Popperian challenge lies in a new model of rationality rid of all justificationist ingredients*: Popper *divorces reason from positive reasons, truth from certainty, argument from belief*. Reason has negative powers and is not a tool of edification but of a purge; as Miller puts it, "we deduce consequences from what we know not in order to … consolidate our knowledge, but to liquidate it" (1994, 111). Miller argues that good reasons not only do not exist, but are neither necessary nor desirable (1994, chap. 3). In Popper's words, "the problem of giving positive justifying reasons" should be replaced by "the problem of *critically* discussing hypotheses" (Popper 1979, 22–23). Popper's conception of reason is deliberately destructive as reason can only undermine our beliefs and remind us of our acute fallibility. Similar to Popper´s appeal to embrace life in uncertainty this situation, too, should be viewed not as paralyzing but exciting. This approach encourages bold guessing, conflict and disagreements, and enhances the culture of open debate and tolerance.

Popper´s new concept of rationality certainly overturns the traditional values ascribed to reason; it is original, provocative and demands courage. David Miller presents this shift as relatively straightforward; we have been hooked on justification, he says, but "it is high time philosophers kicked the habit. Cold turkey is recommended" (Miller 1994, 49). The cold turkey treatment is brutal—in this case, I suspect, too brutal—overlooking the fact that, as Popper himself observes, we have

---

[8] Popper gives Einstein as an example of such a critical attitude, specifically his prediction of the redshift in the spectrum of the satellites of Sirius; if this prediction was wrong, Einstein claimed, the general theory of relativity would be untenable.

an instinctive need for reassurance, an innate (and dogmatic) tendency to look for justification of our expectations. We may eliminate justification through a rational decision yet at heart we still yearn for security. I believe that both Popper and Miller may have underestimated this rift between reason and nature, in which the Humean dilemma resurfaces on a different level: the imperative of reason opposes our natural (but irrational) attachment to a safe, predictable and stable world.

It would be uplifting to imagine that people will increasingly appreciate the benefits of erring, giving up their own hypotheses without a feeling of disappointment and failure and being thrilled to discover new problems instead. As I discuss elsewhere (Parusniková 2004), it is more realistic to see criticism (especially self-criticism) undermined by the natural dogmatic force. Yet such dualism between reason and nature would strengthen Popper's case, reinforcing the need for the methodological maxim of falsification.

But then again, Popper had great faith in reason. He called himself the last laggard of the Enlightenment and, in tune with the spirit of the *Spätaufklärung* of the Vienna Circle, he assigned reason a progressive mission in knowledge and a liberating mission in society. He saw the development of mankind as a process of maturing towards critical rationality; and as the first impetus was given by the development of the argumentative form of language, stimulated further by the emergence of writing, so may the next phase be typical of promoting our rational capacities and internalizing the critical imperative. Of course the future is open, as Popper says; the dangers of irrationalism and—in the social sphere—of various forms of totalitarianism always loom. Yet his philosophical concept of *ratio negativa* brings an optimistic message of progress and freedom.

Acknowledgements I would like to thank David Miller for helpful comments.

References

Agassi, J., Jarvie, I.C. (eds.): Rationality: the critical view. Martinus Nijhoff, Dordrecht (1987)

Fuller, S.: Kuhn vs Popper. Icon Books, Cambridge (2003)

Hume, D.: A treatise of human nature. In: Ed. L.A. Selby-Bigge, revised by P.H. Nidditch. Second edition. Clarendon Press, Oxford (1981[1739])

Hume, D.: An enquiry concerning human understanding. In: Ed. L.A. Selby-Bigge, revised P.H. Nidditch. Third edition. Clarendon Press, Oxford (2005[1748])

Kuhn, T.S.: The structure of scientific revolutions. Chicago University Press, Chicago (1970)

Kuhn, T.S.: Logic of discovery or psychology of research? In: Lakatos, I., Musgrave, A. (eds.) Criticism and the Growth of Knowledge, pp. 1–25. Cambridge University Press, Cambridge (1982)

Miller, D.: Critical rationalism: a restatement and defence. Open Court, Chicago and La Salle, Illinois (1994)

Newton-Smith, W.H.: The rationality of science. Routledge and Kegan Paul (1981)

O'Hear, A.: Karl Popper. Routledge and Kegan Paul (1980)

Parusniková, Z.: Horror Dogmatis. In: Bar-Am, N., Gattei, S. (eds.) Encouraging Openness: Essays for Joseph Agassi on the Occasion of his 90th Birthday. Boston Studies in the Philosophy and History of Science. Springer Verlag, pp. 27–39 (2017)

Parusniková, Z.: Two cheers for Karl Popper. In: O´Hear, A. (ed.) Karl Popper: Critical Assessments of Leading Philosophers. Routledge, London, pp. 79–99 (2004)

Popper, K.R.: The logic of scientific discovery. Harper and Row, New York (1965 [1959])

Popper, K.R.: Objective knowledge: an evolutionary approach. Clarendon Press, Oxford (1979 [1972])

Popper, K.R.: Unended quest. Fontana, Glasgow (1982a [1974])

Popper, K.R.: Normal science and its dangers. In: Lakatos, I., Musgrave, A. (eds.) Criticism and the Growth of Knowledge, pp. 51–59. Cambridge University Press, Cambridge (1982b)

Popper, K.R.: Conjectures and refutations: the growth of scientific knowledge. Routledge and Kegan Paul, London (1985 [1963])

Popper, K.R.: Realism and the aim of science: postscript to the logic of scientific discovery. In: Bartley III, W.W. (ed.). Hutchinson, London (1988 [1983])

Russell, B.A.W.: The problems of philosophy. Williams and Norgate, London (1912)

Russell, B.A.W.: A history of western philosophy. George Allen & Unwin Ltd. Repaginated Classics edition (2004). Routledge, Abingdon (1946)

Salmon, W.C.: The justification of inductive rules of inference. In: Lakatos, I. (ed.) The problem of inductive logic, pp. 24–43. North-Holland Publishing, Amsterdam (1968)

Watkins, J.: Against 'normal science.' In: Lakatos, I., Musgrave, A. (eds.) Criticism and the Growth of Knowledge, pp. 25–39. Cambridge University Press, Cambridge (1982)

Watkins, J.: Science and scepticism. Princeton University Press, Princeton (1984)

Worrall, J.: Revolution in permanence: Popper on theory-change in science. In: O'Hear, A. (ed.) Karl Popper: philosophy and problems, pp. 75–103. Cambridge University Press, Cambridge (1996)

Zuzana Parusniková is a Research Professor in the Institute of Philosophy, Czech Academy of Sciences in Prague. After finishing her PhD at the Charles University in Prague she held the Lakatos Fellowship at LSE in London and was awarded the Karl Popper Stiftung at the University of Cambridge, UK. Her research focuses mainly on two fields: on the philosophy of science, most importantly on critical rationalism, and on early modern philosophy, especially David Hume. Among others she published a monograph on Karl Popper (in Czech), co-edited (with Robert S. Cohen) and contributed to Rethinking Popper (Springer 2009) and is the author of David Hume, Sceptic (SpringerBriefs 2016). E-mail: parusnikova@flu.cas.cz

Physics and Cosmology

Popper and the Quantum Controversy

Flavio Del Santo and Olival Freire Jr.

1 Introduction

It is almost a truism to say that the philosophy of science systematized by Karl Popper (1902–1994) was heavily influenced by the intellectual landscape of physics. Indeed, falsifiability as a criterion to discriminate science from other forms of knowledge was largely indebted to Einstein's predictions drawn from his general theory of relativity. At the same time, Popper's falsificationism left a deep and long-lasting mark on the way physicists perceived their common practice. However, Popper's contribution to the philosophy of quantum physics and its influence among practitioners has been long overlooked and only in recent years has this issue gathered some historiographical attention (Freire 2004; Shields 2012; Howard 2012; Del Santo 2018, 2019, 2020).

Popper's contributions to foundations of quantum mechanics can be divided into three main periods. As early as 1934 he conceived a thought experiment which allowed him to confront the founding fathers of quantum physics such as Albert Einstein, Werner Heisenberg and Niels Bohr. However, this proposal turned out to be mistaken and this accident led Popper away from the quantum controversy for several years.

The second period of Popper's involvement in the debate over quantum foundations spans 1950s–1960s, when he formulated a new interpretation of probability

---

F. Del Santo
Faculty of Physics, University of Vienna, Vienna, Austria

Institute for Quantum Optics, Quantum Information, Austrian Academy of Science, Vienna, Austria
e-mail: delsantoflavio@gmail.com

O. Freire Jr. (✉)
Instituto de Física, Universidade Federal da Bahia, Salvador, Brazil
e-mail: freirejr@ufba.br

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_2

and at the same time of quantum mechanics—the propensity interpretation—gathering the support of several important physicists including David Bohm and Louis de Broglie. In particular, at the end of 1960s, Popper published two influential papers—"Quantum Mechanics without the Observer" (Popper 1967; see Del Santo 2019) and "Birkhoff and von Neumann's interpretation of Quantum Mechanics" (Popper 1968; see Del Santo 2020)—which allowed him to cross the disciplinary borders and become a full-fledged member of the physics community concerned with quantum foundations in the following years.

Finally, in the 1980s, Popper gave another remarkable contribution to foundations of quantum mechanics, publishing a comprehensive volume on *Quantum Theory and the Schism in Physics* (Popper and Bartley 1982). Here he also proposed a new version of the Einstein-Podolsky-Rosen (EPR) thought experiment alleged to put to the test Heisenberg's uncertainty principle, and the whole Copenhagen interpretation along with it. At the time, Popper was able to count on the strong support of physicists such as Jean-Pierre Vigier and Franco Selleri, who were harsh critics of the Copenhagen interpretation of quantum physics (see Freire 2004; Del Santo 2018).

Initially conceived as a thought experiment, Popper's EPR-like proposal eventually found its way, at the end of the twentieth century, onto lab benches thanks to Yanhua Shih. The interpretation of this experiment triggered a lasting debate that survived Popper himself, as Kim and Shih's results (Kim and Shih 1999) were disconcerting and triggered a real stir, which still deserves historical investigation. In hindsight, we may say that much of the debate was related to a poor understanding, even among physicists, about "entangled" pairs of photons. Indeed, the issue was reviewed, a decade later, by Tabish Qureshi, whose resolution of this issue states: "[Popper's] experiment, by its very nature, cannot be decisive about Popper's test of the Copenhagen interpretation, a point missed by both Popper and the defenders of the Copenhagen interpretation" (Qureshi 2012). Qureshi concludes that "Popper's experiment has proved to be useful in understanding what quantum correlations are, and more importantly, what they are not."

In a nutshell, Popper's ideas on the foundations of quantum mechanics may be summarized as being based on the assumptions of both *realism* and *indeterminism*. Indeed, he fully accepted the intrinsically probabilistic nature of physical processes (actually also at the classical level) and, motivated by this, he suggested his propensity interpretation as an interpretation of probability which later was converted into an interpretation of quantum mechanics. Without any attachment to determinism, Popper criticized the introduction of subjectivist approaches in this scientific domain, aligning himself with the realist position in the quantum controversy, while harshly criticizing the widespread Copenhagen (or orthodox) interpretation (see Freire 2015).

In this chapter we present a chronologically organized overview of Popper's concerns with quantum mechanics, and, as an epilogue, we summarize the debates about the experiment he had suggested, and assess the resonance of Popper's indeterministic view on current research in (quantum) physics.

2 Popper and Quantum Mechanics

2.1 Popper's Early Concerns with Quantum Theory (1934)

Remarkably, Popper's engagement in the debate over the foundations of quantum mechanics dates back to the early days of the theory, and eventually lasted for the rest of Popper's life. As early as 1934, Popper conceived a thought experiment which was devised to advocate a statistical interpretation of the Heisenberg's uncertainty relations, as opposed to a fundamental limitation to the determinacy of conjugated variables in a single quantum system. In Popper's words, this thought experiment turned out to be "a gross mistake for which [he had] been deeply sorry and ashamed of ever since" (Popper and Bartley 1982); and yet this accident allowed him to confront the founding fathers of quantum physics—among whom figure the names of Einstein, Heisenberg and Bohr—and it possibly even had some influences on Einstein in his subsequent development of the EPR paradox (see Jammer 1974, 178). However, Popper's mistake (together with the tragic historical events that shook Europe in the 1930s and1940s) led Popper away from the quantum controversy for over a decade.

It was only in 1948 that Popper returned to think about problems of quantum foundations, mostly thanks to the encouragement of his friend, the Austrian physicist Arthur March (see Popper 1976, 106). It was around the same time that Popper's ideas on indeterminism began to take shape: In November 1948, he gave a first talk at the British Society for the History of Science on "Indeterminism in Quantum Physics and in Classical Physics"; he then presented the same topic in a course of lectures he held at Harvard University, and again in 1950 in Princeton in front of Einstein and Bohr (see Del Santo 2019). These ideas appeared in print, too, soon after (Popper 1951). Popper proposed the novel view that both classical and quantum physics can (and ought) to be interpreted indeterministically (as we will see in Sect. 3.2, these ideas had an influence on similar recent developments).

Through the radicalization of his stance regarding indeterminism in physics starting from 1953 Popper developed the conceptual tool of "propensities", namely objective intrinsic probabilities that determine the tendency for a certain physical process to happen in a genuine indeterministic way. This innovative idea brings together physical (indeterministic) processes and mathematical probabilities in a natural way. Popper, in fact, proposed "that probabilities must be 'physically real'—that they must be physical propensities, abstract relational properties of the physical situation" (Popper 1959). He publicly presented this interpretation of probabilities for the first time in April 1957, at the "Ninth Symposium of the Colston Research Society in Bristol", publishing two papers on this topic (Popper 1957, 1959).

As a matter of fact, it should be noted that throughout the 1950s, while being explicitly physically motivated, the propensity interpretation remained no more than a formal interpretation of the calculus of probability and its resonance among physicists was negligible at the time. As we shall see, Popper's role in the quantum debate was to be drastically boosted in the following decade, when his propensity interpretation became an actual comprehensive attempt to interpret the quantum theory.

2.2 The Turning Point: From Philosophy to Physics (Ca. 1967–1968)

Before moving forward to discuss Popper's further contributions to the quantum controversy, a clarification of a more sociological nature seems due. In fact, as argued in detail in (Del Santo 2019), there is good evidence to maintain that up until 1960s none of the aforementioned efforts that Popper made in the field of foundations of quantum mechanics had almost any influence in the community of physicists (besides the mistaken thought experiment of 1934). As a matter of fact, in those years Popper was a reference point for some physicists, notably Alfred Landé, David Bohm and Hermann Bondi, with some interest in philosophy. He helped them to network and even publish philosophical papers, but the resonance of his own ideas among physicists remained scarce. It was only in the mid-1960s, thanks to new acquaintanceships with physicists who were active also in the community of philosophers of science—in particular Wolfgang Yourgrau and Mario Bunge—that Popper's ideas started to become influential among physicists. This led to the real turning point of Popper's role in the quantum debate, namely the publication of two papers, "Quantum Mechanics without the Observer" (Popper 1967; see also Del Santo 2019) and "Birkhoff and von Neumann's interpretation of Quantum Mechanics" (Popper 1968; see also Del Santo 2020), which projected Popper into discussions with several physicists active in the foundations of quantum theory.

In the latter of these two papers, Popper claimed that an extremely influential proposal by Garrett Birkhoff and John von Neumann (Birkhoff and von Neumann 1936)—which initiated the subfield known as the "logic of quantum mechanics" (LQM)—was formally flawed. LQM is an axiomatic approach to quantum theory that describes physical systems in terms of "yes-no questions" (or empirical propositions) and investigates the algebraic structures of the logical connectives that relate them, which are compatible with the observed phenomenology. Now, in classical physics, the state of a system is a mathematical point in phase-space, thus any yes-no question, e.g. "is the position of a particle in the interval [0,1]?", has a fully determined truth value at each time; and so it is the conjunction and the disjunction of any two propositions. It can be shown that the empirical propositions of classical physics are compatible with Boolean algebra. On the other hand, in their pioneering work, Birkhoff and von Neumann showed that Boolean logic is incompatible with the phenomenology of quantum mechanics, due to Heisenberg's uncertainty principle.[1] In the late 1960s, LQM was experiencing a revival, specially due to the "school of Geneva" which gathered around the figure of Joseph-Maria Jauch. It was this renewed interest that led Popper to write a critical paper against the whole approach, which was rooted in the standard interpretation of the uncertainty relations (considered by

---

[1] Technically speaking, what fails in quantum mechanics is the distributive law (which is one of the properties that characterizes Boolean algebra) for empirical propositions, due to the existence of incompatible observables (i.e. not commuting operators); see (Del Santo 2020) and references therein for further details.

Popper a crucial part of the Copenhagen interpretation), which Popper had already tried to dismantle in the 1930s.

It would be impossible to analyze here Popper's criticisms in detail. They are rather technical, but, as a matter of fact, they turned out to be mostly based on misconceptions, as also later acknowledged by Popper himself (who, in fact, did not reproduce any of these arguments in his book on the philosophy of quantum theory: Popper and Bartley 1982). Nevertheless, from the historiographical point of view, Popper's critique of LQM is an interesting case. In fact, the reputation of Popper as a philosopher and of Birkhoff and von Neumann as mathematicians, together with the distinction of the journal *Nature* on whose pages the paper appeared, made historians wonder why this incident did not trigger a broad debate. Indeed, only recently has one of the present authors (FDS) reconstructed the genesis of Popper's efforts against LQM in detail, and has shown that not only the short paper in *Nature* (Popper 1968) was merely one of five manuscripts (the others remained unpublished but are now partly retrieved, see Del Santo 2020), but also that this debate did happen albeit in the form of private correspondence. Indeed, Popper had a sustained epistolary exchange with many of the protagonists of the new LQM and Jauch in particular. The latter even went as far as accusing Popper of collusion, when a critical comment by Arlan Ramsay and James C. T. Pool was rejected by *Nature*. He wrote to Popper: "You have published in a widely read periodical criticisms of an important paper, which you have certainly misunderstood. [...] You realize of course that the entire scientific progress depends on the possibility of free exchange of scientific information and criticism. [...] Did you not say yourself in the "Open Society and its Enemies" the spirit of science is criticism. If you believe that, I suggest that you send the enclosed copy of the manuscript by Ramsay and Pool to Nature with your personal request that it be published." (Letter to Popper on February 24th, 1969. Reproduced from Del Santo 2020). This triggered Popper's outrage, who replied: "I do not see what can give you the right to suppose that there is a need to remind me of this; or what your remark may mean unless you wish to accuse of dishonesty." (letter from Popper to Jauch on February 28th, 1969. Reproduced from Del Santo 2020). Although Popper solicited the publication, this critical comment never appeared in print but, thanks to the interaction with the mathematician Simon Kochen, Popper eventually was persuaded that his criticisms were based on a misunderstanding of the original paper of Birkhoff and von Neumann (which admittedly had some ambiguous definitions). It ought to be stressed, however, that this period of intense debate with a number of physicists and mathematicians—besides the aforementioned Jauch, Ramsay, Pool, and Kochen, also David Finkelstein, Abner Shimony, de Broglie—clearly helped pave the way for Popper's entrance into the community of quantum foundations in the following years.

However, the publication that most of all broke the ice for Popper's interaction with quantum physicists was "Quantum Mechanics without the Observer'" (QMwO), which even today arouses some theoretical interest, besides its historical importance. In the paper Popper presents a physical interpretation of quantum mechanics, and propensities are no longer merely an interpretation of probability from which one could indirectly infer an interpretation of quantum mechanics. He presents what a little later became well known as the statistical or ensemble interpretation of quantum mechanics. Indeed, in 1970, Leslie Ballentine christened "The Statistical Interpretation of Quantum Mechanics," as the interpretation "according to which a pure state (and hence also a general state) provides a description of certain statistical properties of an ensemble of similarly prepared systems, but need not provide a complete description of an individual system," and attributed it to Einstein, Popper, and Blokhintsev (Ballentine 1970, 360). Twenty years later Dipankar Home and M. A. B. Whitaker reviewed the statistical interpretation, rechristened it as the "ensemble interpretation," and related it to the diverse interpretations of probabilities. Popper is presented, again, as an advocate of such an interpretation. Thus, it is beyond doubt that with QMwO, Popper entered the physics scene as a proponent of a physical interpretation of quantum mechanics.

Popper presented his views in the schematic form of 13 main theses. For him, quantum mechanics is a theory about statistical problems, such as black-body radiation, and not about atomic stability (1st thesis); "statistical questions demand, essentially, statistical answers, thus quantum mechanics must be, essentially, a statistical theory" (2nd thesis, p. 170); and, this way, there is no "no lack of knowledge, which allowed the intrusion of the observer," (3rd thesis). The following thesis is about what Popper called the "great quantum muddle," a view about the object of statistical distributions. According to Popper, statistical distribution functions "may be looked upon as *a property characterizing the sample space*," [as] "it is *not* a physical property characteristic of the *events* […]; still less is it a property of the *elements*." Thus, "the great quantum muddle consists in taking a distribution function, i.e. a statistical measure function characterizing some *sample space* […], and treating it as *a physical property of the elements of the population.* It *is* a muddle: the sample space has hardly anything to do with the elements." The philosophical sophistication of these remarks did not pass by commentators unnoticed (Home and Whitaker 1992, 280).

Popper continued by stating, in the 5th and 6th theses, that formulae such as Heisenberg relations are statistical scatter relations (as he had already maintained since 1934). He presented an approximate derivation of these relations departing from equations of classical physics, optics for instance. In these derivations Popper assumed quantum systems as always having well defined properties such as positions and momenta previous to the measurement. It is also noticeable that Popper did not strictly appeal to the quantum mechanics mathematical formalism. In the 8th thesis Popper suggested that while quantum mechanics is a statistical theory, it is applicable to singular systems. But these systems are not things such as electrons. Indeed, for Popper, probability statements are "statements about some measure of a property (a physical property, comparable to symmetry or asymmetry) of the whole experimental arrangement; a measure, more precisely, of a virtual frequency" (p. 32). Thus "propensities are properties of neither particles nor photons nor electrons nor pennies. They are properties of the repeatable experimental arrangement - physical and concrete, in so far as they may be statistically tested" (p. 32). Some readers would see a rapprochement to Bohr's position here; but far from this, Popper's position was grounded on an explicit defense of realism, as we are going to see. Before this, Popper's last thesis was the statement that he was not concerned with the quantum indeterminism. Reasons for this go beyond the consideration of the quantum case and encompass the whole of physics; for him, "both classical physics and quantum physics are indeterministic" (p. 40).

Popper's realism was larger than the assumption, so common among physicists, of the existence of a reality independent of the existence of an observer. He called for the distinction between theories and concepts, and assumed that theories are statements about the world. Thus, it would be wrong to take physical theories as "conceptual systems" or "conceptual frameworks." He acknowledged that "it is true that we cannot construct theories without using words or, if the term is preferred, 'concepts'." But insisted "it is most important to distinguish between statements and words, and between theories and concepts." While grounded on the logical distinction between words and statements, Popper had a precise target in the world of quantum physics. Indeed, he criticized all the physicists who, following Ernst Mach, defended the view that physical theories are mostly about concepts, calling this is an instrumentalistic view of science. Finally, he criticized Niels Bohr for adhering to Mach's position, and concluded by criticizing the idea of complementarity between the wave picture and the particle picture as a tenet of quantum theory, as pictures could not be essential parts of a physical theory.[2]

QMwO, which was urged by Bunge and published in a volume edited by him, was perhaps the Popper's first paper (since the 1930s) targeting an audience of physicists. And indeed, it soon started to bear fruit: Bohm was the first to write to Popper praising his propensities[3]: "I feel that what you have to say about propensities makes a genuine contribution to clarifying the issues that you discuss." (Letter from Bohm to Popper, on March, 3rd 1967. Reproduced from Del Santo 2019). Also Landé and Bondi, who both were friends with Popper, reacted positively to QMwO. However, what is remarkable is that the resonance of Popper's paper transgressed his usual circle of acquaintances. Among others, Bartel L. van der Waerden—a former pupil of Emmy Noether in Göttingen a close collaborator of Heisenberg in Leipzig—wrote to Popper concerning QMwO: "I fully agree with your 13 theses, and I feel it was very good you expounded them so clearly. I also agree with your propensity interpretation of probability. […] I feel my ideas are in perfect accordance with your theses" (letter from van der Waerden to Popper, on October 19th, 1968. Reproduced from Del Santo 2019). Finally, also the French Nobel laureate and founding father of quantum theory, Louis de Broglie, wrote to Popper: "I noticed with great pleasure that your ideas are very close to mine". (de Broglie to Popper, on March 4th, 1969. PA 96/7. Reproduced from Del Santo 2019).

---

[2] All citations from Popper (1967, 11–14).

[3] It should be stressed that Bohm was present at the first symposium in 1957, when Popper's propensities were first presented. Moreover, he had been regularly in touch with Popper for a decade since then, but it seems that he was not aware of propensities yet. This corroborates our thesis that up until QMwO physicists, even those close to Popper, paid little attention to Popper's ideas related to quantum foundations.

Popper's QMwO also received several rebuttals: Jeffrey Bub, a former student of Bohm in London, rejected Popper's propensities (Bub 1972; see Jammer 1974, 452–453), showing that Popper's interpretation of quantum theory in terms of propensities is problematic because it is an interpretation of a still Boolean probability calculus which is not compatible with quantum probability. Such a criticism was similar to that levelled by Paul Feyerabend, who published a vitriolic rebuttal (Feyerabend 1968) of Popper's QMwO, in the course of his vaster critique of Popper's ideas in those years.[4]

In point of fact, Popper's role in the quantum debate changed drastically after the late 1960s, and his influence among physicists became appreciable. In particular—more than likely through the common friendship of Bohm—Popper got to know Jean-Pierre Vigier, a French physicist, pupil of de Broglie, who had contributed a great deal to the realistic program in quantum foundations. He was to become a valuable ally for Popper, and it is mostly thanks to his encouragement that Popper entered his last period of activity on quantum foundations, this time fully within the community of physicists.[5]

2.3 The Mature View: Popper's Experiment (The 1980s)

Popper entered the 1980s, well into his eighties, with a new turn in his intellectual life. This was related to the space he opened to the research in quantum mechanics. Thanks to the regular interaction with Vigier, he had been thinking about new experimental proposals to confirm the realistic interpretation of quantum theory, while ruling out the Copenhagen one. And, indeed, in June 1980, Popper devised a variant of the EPR experiment—currently known as Popper's experiment—to enlighten the foundations of quantum mechanics (see Del Santo 2018). This was, however, published only two years later in the long awaited three volumes of the *Postscript* to the *Logic of scientific Discovery*, which were fully dedicated to the philosophy of science. Most of the content of the volumes in this series had already been written in the late 1950s but had not been published due to Popper's health issues and other incidental reasons. The third volume, entirely dedicated to the interpretation of quantum mechanics, was meaningfully entitled *Quantum Theory and the Schism in Physics* (Popper and Bartley 1982). Moreover, Popper strengthened his engagement with physicists more than ever before: he authored papers published in physics journals, established lasting intellectual relationships with some of the protagonists of the quantum debate (notably, besides Vigier, with Franco Selleri and the initiators of the

---

[4] Sect. 3.4 of (Del Santo 2019) is devoted to the debate between Popper and Feyerabend triggered by the publication of QMwO. A dedicated paper is in preparation: Del Santo, F. "Beyond method: The diatribe between Feyerabend and Popper over the interpretation of quantum mechanics", to appear in a special issue edited by M. Stuart and J. Shaw on "Feyerabend and the History and Philosophy of Physics" in *Studies in History and Philosophy of Modern Physics.*

[5] It should also be remarked that Popper has been among of the first authors to respond in print to the pivotal result of Bell's theorem, at a time when it was completely overlooked (Popper 1971).

revival of foundations of quantum mechanics in Italy) and he was also appointed member of advisory committees of international physics conferences on quantum foundations. In this way, Popper lent his intellectual and social prestige to the cause of the "quantum dissidents", namely, those physicists fully dedicated to the development of research on the quantum foundations.[6]

This stage of Popper's activities was to leave a legacy which would continue to be fruitful years after Popper's passing. In the late 1990s, the physicists Yanhua Shih and Yoon-Ho Kim performed the experiment Popper had suggested and this still arouses debate today. And yet, Popper's criticisms towards determinism in the foundations of quantum and classical mechanics would resonate with later physics research in these domains.

Thus, in the 1980s, based on these ideas, Popper was ready to have stronger interaction with physicists than he had had so far. He collaborated with Vigier and the Italian young physicist Augusto Garuccio (a pupil of Selleri) to suggest a new experiment to test the existence of the empty waves Louis de Broglie had suggested in the mid-1920s. Furthermore, Popper presented the aforementioned modified version of the EPR experiment in order to invalidate Heisenberg's relation, as interpreted according the orthodox manner, and tried to persuade physicists to perform it. Indeed, this experiment was the true novelty in *Quantum Theory and the Schism in Physics*. The background and history of these experiments have been narrated in detail by one of the authors (FDS), so just to summarize here.

In 1952, David Bohm had suggested his interpretation in terms of hidden-variables (Bohm 1952), without previous knowledge of Louis de Broglie's earlier works (see Freire 2019). In fact, Bohm had developed what we call the pilot wave model in order to overcome some criticisms Wolfgang Pauli had addressed to this work. De Broglie then came back to his earlier ideas, but instead of defending the pilot wave model he had presented at the Solvay Council in 1927, he resumed ideas he had published before that. De Broglie called these ideas, in fact a model and a research program, the double solution. De Broglie meant to represent quantum systems by two equations, the first describing a wave guiding a particle (similar to solutions of the Schrödinger's equation), and the second, a nonlinear and so far unknown equation representing the particle itself. During the 1950s, neither Bohm nor de Broglie tried to test their ideas to the usual interpretation on the lab benches. Indeed, they emphasized their empirical equivalence. In the late 1970s, under the influence of the experiments on Bell's theorem, which opposed quantum mechanics to theories based on the assumption of local realism, Vigier began to look for experiments to test de Broglie's double solution. Until then, it had also been known as the "empty wave" because in the two-slits experiment, it suggests the particle passes through one of the slits and the wave passes through both slits, thus there is an empty wave through the slit where the particle did not pass. In the early 1980s, Vigier and Garuccio were suggesting a modified Mandel-Pfleegor experiment to test the empty wave proposal and invited Popper to co-author the papers with such a proposal. Popper accepted for he was sympathetic with the empty wave proposal which was, in fact, compatible with

---

[6] Popper and Bartley (1982), Del Santo (2018), Freire Junior (2015).

the statistical interpretation Popper was advocating: On the one hand, the statistical interpretation was silent about physical models governing individual systems, and, on the other hand, the empty wave was silent about the equation governing the particle. If Vigier had suggested an experiment to test the Bohm-de Broglie pilot wave, instead of the double-solution (empty wave) idea, we may wonder Popper would have difficulty joining the enterprise as the pilot wave was deterministic and Popper had maintained that both quantum and classical physics are indeterministic (see above). The suggested experiment, however, did not materialize.[7]

Regardless of the joint papers on the empty wave (but admittedly stimulated by discussion with Vigier), Popper presented in his *Quantum Theory and the Schism in Physics,* in Sect. 9 of the preface of this book, what is nowadays known as Popper's experiment (PE). It was presented as "a simple experiment which may be regarded as an extension of the Einstein-Podolsky-Rosen argument." Indeed, despite the apparent similarity with the EPR experiment, there are substantial differences between them. While EPR exhibited and refused the quantum nonlocality, PE was conceived to reveal limitations in Heisenberg's uncertainty relations, or, to be more precise, in a certain interpretation of Heisenberg's relations. Popper assumed a pair of quantum particles were created and emitted (coaxially) in opposite directions, then he suggested a positronium as the source for a pair of photons, each one passing through a slit, A and B, and which were detected on screens behind the slits. The slits may be moved. Popper's drawing illustrates the idea (Fig. 1).

From the width of the slit A we may obtain the scattering of the position of the particle on the right-hand side in the y direction, thus Δq_y and, through Heisenberg's relations, the uncertainty in the momentum Δp_y. As the particles were emitted in opposite directions, one can infer the same Δq_y for the left particle, thus also its Δp_y. The p scatter may be measured through the angle of the detectors being fired at the screens. Now let us follow Popper's argument in his own words (Popper and Bartley 1982, Preface, Sect. 9):

> Now we make the slit at A very small and the slit at B very wide. [...] we have measured q_y for both particles (the one passing through A and the one passing through B) with the precision Δq_y of the slit at A, since we can now calculate the y-coordinate of the particle that passes through B with approximately the same precision, even though its slit is wide open. We thus obtain fairly precise '*knowledge*' about the q_y position of this particle—we have 'measured' its y position indirectly. And since it is, according to the Copenhagen interpretation, *our knowledge* which is described by the theory—and especially by the Heisenberg relations—we should expect that the momentum p_y of the beam that passes through B scatters as much as that of the beam that passes through A, even though the slit at A is much narrower than the widely opened slit at B. Now the scatter can, in principle, be tested with the help of the counters. If the Copenhagen interpretation is correct, then such counters on the far side of B that are indicative of a wide scatter (and of a narrow slit) should now count coincidences: counters that did not count any particles before the slit at A was narrowed.

Popper concludes, "to sum up: if the Copenhagen interpretation is correct, then any increase in the precision of our *mere knowledge* of the position q_y of the particles

---

[7] On the Solvay council, see Bacciagaluppi and Valentini (2009); on de Broglie's story of his interpretations, see de Broglie (1956). The two papers co-authored by Popper are Garuccio et al. (1981a, b).

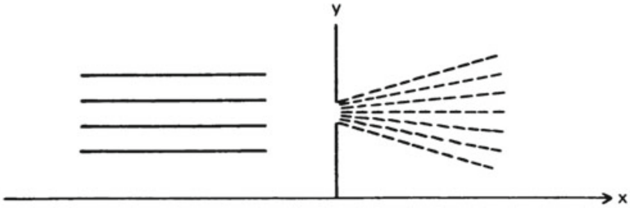


Figure 1

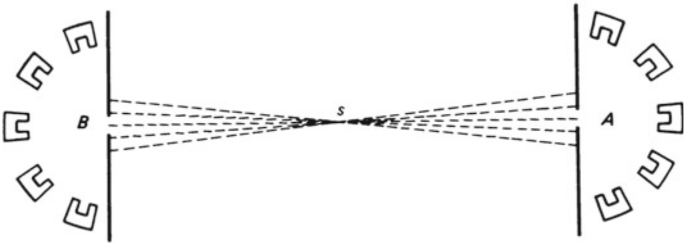


Fig. 1 Reproduction of Figs. 1 and 2 of Popper's postscript to the logic of scientific research (1982, 17 and 28, respectively), portraying Popper's EPR-like experiment. With permission of University of Klagenfurt/Karl Popper library. All rights reserved

going to the left should increase their scatter; and this prediction should be testable," and follows expressing his own expectations about the suggested experiment and its distinct implications.

Having published his suggested experiment in a book, not the place physicists would usually look for new ideas, Popper began his peregrination to convince the scientific community about the relevance and feasibility of the experiment. Towards this goal, connection with Selleri was instrumental. Introduced to the Italian physicist by Vigier, Popper was invited to attend a conference Selleri was organizing in Bari in 1983 to which Selleri had invited a few physicists just to listen Popper and discuss his experiment. The debates gathered people such as Marcello Cini, Francesco De Martini, Karl Kraus, Trevor Marshall, Helmut Rauch, Gino Tarozzi, C. Robinson, J. Six, in addition to Selleri and Vigier themselves. Most of the debate concerned the feasibility, more particularly how to obtain from a point source, a collinear pair of photons compatible with the width of the slits in suggested experiment. Noticeably, nobody noticed that the joint detection of the particles in the two screens implied describing them through an entangled state of the two particles, thus they could not be described as independent systems. After the Bari conference, Popper kept up the discussion of his suggested experiments but ultimately nothing materialized to bring the thought experiment to the world of real ones.[8]

Popper's activities from the late 1960s to the 1980s had an intellectual influence on the research on the foundations of quantum mechanics beyond the influence of his philosophical and scientific ideas. He was already an influential philosopher in the public sphere and brought this prestige to the "quantum dissidents", i.e. the small number of physicists who were challenging the dominant views on the interpretation of quantum mechanics and badly in need of support at those times. A letter from Selleri to him encapsulates the debt the quantum dissidents had to the Austrian philosopher: "This is the real strong idea [realism] that we have in common and I am always very grateful for the great battle you fought and you fight against the idealistic conceptions of the Copenhagen school. You gave us some water in which we can now try to swim."[9]

3 Epilogue: Popper's Legacy in Quantum Physics

3.1 Kim and Shih, and the Real Popper Experiment

The experiment suggested by Popper had an afterlife that would have surprised and pleased him, but he was no longer alive to follow its developments. The subject was resumed in the mid-1990s, when Garuccio explained PE to the Sino-American physicist Yanhua Shih, at the University of Maryland at Baltimore County. The latter immediately got down to work and carried out the experiment with his colleague Yoon-Ho Kim soon after. Shih had been one of the pioneers in the use of a more efficient source of pairs of entangled photons, parametric down conversion (PDC), for experiments in quantum optics. This technique consists of obtaining a pair of entangled photons thorough the nonlinear interaction of one photon of higher energy with a crystal. Indeed, interferometry experiments using PDC photon pairs were pioneered by the two following teams: Carroll Alley and Yanhua Shih at the University of Maryland and Ruba Ghosh and Leonard Mandel at the University of Rochester (Greenberger et al. 1993, 22). Kim and Shih circumvented the issue of obtaining a pair of collinear particles from a point source through the ingenuity of the use of another technique, that of ghost image with the use of a converging lens. The experiment attracted the attention of Vigier and Garuccio, who contributed to the discussion of the experiment.

Kim and Shih's results (Kim and Shih 1999) were disconcerting and triggered a stir not only because of the results themselves but also because the original paper was refused by two very prestigious journals (*Nature* and *Physical Review Letters*), before

---

[8] Tarozzi and van der Merwe (1985). For the debates on the PE experiment in the early 1980s, see Del Santo (2018, 64–66).

[9] Selleri to Popper, 28 November 1989, in Freire Junior (2004, 124). On the quantum dissidents, see Freire Junior (2015).

being eventually published in *Foundations of Physics*. While this journal is a place where research on the foundations of quantum mechanics is usually found, it is not a mainstream physics journal. Interestingly, the behind the scenes of the publication of this paper deserves historical investigation however, this is beyond of the scope of our paper.[10] From the beginning, the authors warned that they were dealing with entangled photons, which means their state must be described through the formalism of quantum mechanics. However, the text did not emphasize this in the introduction of the paper but only in its development. The paper is titled "Experimental Realization of Popper's Experiment: Violation of the Uncertainty Principle?". The abstract states (Kim and Shih 1999):

> An entangled pair of photons (1 and 2) are emitted in opposite directions. A narrow slit is placed in the path of photon 1 to provide the precise knowledge of its position on the y-axis and this also determines the precise y-position of its twin, photon 2, due to quantum entanglement. Is photon 2 going to experience a greater uncertainty in momentum, that is, a greater Δp_y because of the precise knowledge of its position y ? The experimental data show $\Delta y \, \Delta p_y < h$ for photon 2. Can this recent realization of the thought experiment of Karl Popper signal a violation of the uncertainty principle?

The paper follows stating, "it is astonishing to see that the experimental results agree with Popper's prediction." Still, "through quantum entanglement one may learn the precise knowledge of a photon's position and would therefore expect a greater uncertainty in its momentum under the usual Copenhagen interpretation of the uncertainty relations. However, the measurement shows that the momentum does not experience a corresponding increase of uncertainty. Is this a violation of the uncertainty principle?" Only at this point the subject is indeed explained (Kim and Shih 1999, 1850), "as a matter of fact, one should not be surprised with the experimental result and should not consider this question as a new challenge. Similar results have been demonstrated in EPR type of experiments and the same question has been asked in EPR's 1935 paper." One year later, Shih and Kim (2000) were more precise in their wording, saying right in the abstract "the experimental data show that there appears to be a violation of the uncertainty principle. This is, however as we shall argue in this paper, only an illusion provided that we take the teachings of quantum mechanics seriously."

In hindsight, we may say that much of the debate was related to a poor understanding, even among physicists, about the hierarchy among concepts such as "quantum entanglement," strictly quantum concepts on the one hand, and the classical concept of separability which has a limited validity in quantum mechanics, on the other. Entanglement, a word first coined by Erwin Schrödinger, may only be fully understood in the context of the mathematical formalism of quantum mechanics. The issue of the interpretation of PE was reviewed, a decade later, by the Indian physicist Tabish Qureshi. In order to obtain a better comprehension of the issues at stake he translated Popper's experiment, which initially dealt with continuous variables, momentum and position, to a system with discrete variables, which are easier to deal with. It is worth remarking that a similar procedure was done by David Bohm

---

[10] A preliminary discussion on this issue can be found in (Del Santo 2018).

in his 1951 textbook *Quantum Theory* where he introduced the EPR for spin variables instead of momentum and position. Leaving aside the technical details, his main conclusions were that (Qureshi 2012, 28–30) "Kim and Shih correctly implemented Popper's experiment through the innovative use of the converging lens, and the results are in good agreement with the prediction of quantum mechanics and that of the Copenhagen interpretation." The way he uses the term "Copenhagen interpretation," however, should be taken with a grain of salt because for Qureshi it does not include the subjectivist feature Popper would attribute to it. In fact, for Qureshi, "from this point of view, we conclude that the Copenhagen interpretation has been vindicated. It could not have been otherwise, because our theoretical analysis shows that the results are a consequence of the formalism of quantum mechanics, and not of any particular interpretation." Thus what was at stake was just the mathematical formalism of quantum theory. But then should we conclude from this analysis that Popper committed a gross mistake? Far from it. Still following Qureshi, "However, this experiment, by its very nature, cannot be decisive about Popper's test of the Copenhagen interpretation, a point missed by both Popper and the defenders of the Copenhagen interpretation." Indeed, Popper cannot be considered the only responsible for introducing a "flawed assumption." For Qureshi, "all the defenders of Copenhagen interpretation seemed to have the same view, that is why nobody pointed otherwise, and that is the reason why there was so much surprise at the results of Kim and Shih's experiment." Qureshi concluded by noticing that "the problem was that Popper and most of his critics arrived at a wrong conclusion as to what result the experiment would yield. This was simply because no one cared to do a rigorous analysis, but used some commonly understood notions about measurement, which led them to a wrong conclusion. With a lot of theoretical and experimental work in quantum systems behind us, now we are wiser and realize that quantum mechanics is full of such pitfalls." Finally, and endorsing Qureshi's conclusions, "Popper's experiment has proved to be useful in understanding what quantum correlations are, and more importantly, what they are not."

3.2 Popper's Ideas in Contemporary Physics: The Revival of Indeterminism

We would like to conclude this chapter by assessing, as far as possible, the main intellectual marks that Popper's ideas have left on today's fundamental physics.[11] In fact, we cannot but agree with Selleri's words, when stating that "Popper's greatness does

---

[11] Relevant discussions about Popper's contribution to the foundations of physics (besides the many historiographical reconstructions mentioned above) can be found in (Jammer 1991) and (Redhead 1995), respectively written at the end of Popper's career and immediately after his death. More recently, the online Journal *Quanta* devoted its first issue to Popper's philosophy of quantum physics (http://quanta.ws/ojs/index.php/quanta/issue/view/1/showToc), whereas a part (4–B) of (Javie et al. 2006) reassessed some of Popper's work on physics in modern perspective.

not and cannot lie in a series of proposals about the nature and problems of contemporary physics that are ‘all correct’, but rather in an overall framework of ideas of exceptional interest that have filtered through science […]” (Selleri 1990, 351). We will thus not be concerned here with the groundbreaking impact of Popper’s contribution to methodology (i.e. falsificationism) on today’s physics. This is despite the fact that he did deeply change–perhaps more than anyone else–the understanding that most physicists had about their own work and influenced countless research programs (see Del Santo and Cardelli 2019).[12] We will focus here on some outstanding instances of the repercussions that Popper’s ideas on the foundations of physics had within this field of research.

Among Popper’s conceptual contributions, the one that arguably has had the broadest impact on today’s science seems to be the indeterministic nature of physics (even at the classical level). This standpoint is complemented and formalized by the propensity interpretation of probabilities, which allows (objective) causal relations to be maintained even if determinism is refuted. Indeed, “propensity is a form of causality that is weaker than determinism” (Ballentine 2016). Mauricio Suárez, who made important contributions to the propensity program, noted that “Karl Popper’s propensity interpretation of quantum mechanics is surely his most important contribution to the philosophy of physics” (Suárez 2009). Popper’s propensities have become one of the few existing established interpretations of probability in every standard manual on the foundations of probability theory; moreover, it is the only one that allows us to make sense of single-case probabilities (together with the subjective interpretation). However, the legitimacy of propensities as a proper interpretation of probability has been often challenged, notably by P. Humphreys (1985), who noted that if we are to ascribe a causal meaning to probabilities, then the Bayes’ rule for conditional probabilities lacks a meaningful interpretation. To see this, let us consider two events, say, event A “drinking a glass of lemonade”, and event B “outside it is hot”, and let us indicate with $P(A|B)$ the conditional probability of the event A, given event B. Now, if probabilities are interpreted objectively as causal dispositions, it is completely reasonable to attribute a probability to the conditional event “drinking a glass of lemonade, given that it is hot outside”. Yet, it would be foolish to state that being hot outside is causally influenced by someone drinking a lemonade, i.e. express the reversed conditional probability $P(B|A)$. In fact, Bayes’ rule, which follows from the axioms of probability, states that conditional probabilities can be reversed as follows

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

---

[12] We should notice that Popper’s falsificationism does not enjoy large support among philosophers of science today, who have harshly criticized it as a too narrow and naive view. Among Popper’s critics, we ought to mention the greatly influential Thomas Kuhn and Paul Feyerabend, whereas a failure of Popper’s falsificationism in historical perspective has been recently provided by Brush and Segal (2015). Despite this, as noted by Kragh (2013), it can be contented that “Karl Popper’s philosophy of science […] is easily the view of science with the biggest impact on practising scientists”.

where *P(A)* and *P(B)* are the marginal probabilities of observing event *A* and *B*, respectively. More generally, if probabilities are propensities, this means that they express the objective tendency of an effect to happen given a certain cause but, according to Bayes' rule, also the cause would in turn be causally influenced by the effect and this is paradoxical (*Humphrey's paradox*). Thus, Humphreys concluded that "propensities cannot be probabilities" (Humphreys 1985).

Several philosophers have vindicated Popper's propensities as fully fledged probabilities, most prominently Miller (1994). But the kind of solutions of Humphrey's paradox that they have proposed take propensities only to refer to present events, and this requires that propensities depend upon *all* past causal influences, i.e. the whole past light cone. Interestingly, some physicists (more or less aware of this) accepted Humphreys' criticism, thereby developing a theory of propensities which do not satisfy the axioms of probability (e.g. Gisin 1991; Ballentine 2016). This does not make propensities a full-fledged interpretation of probabilities, but allows them to be a useful tool to describe indeterministic physical theories.

In recent years, these ideas have, in fact, gained new impetus among physicists. Ballentine has further developed the original idea of Popper's "that the intrinsic quantum probabilities (calculated from a state vector or density matrix) are most naturally interpreted as quantum propensities." (Ballentine 2016; see also Maxwell 2011). Another novel line of research which arguably has a clear resonance with the Popperian program has been put forward by Nicolas Gisin in a series of papers (Gisin 2019; Del Santo and Gisin 2019; Gisin 2020). In these, he has challenged the foundations of classical physics, showing that the seemingly innocuous standard assumption that physical quantities take values in the real numbers leads to the unphysical possibility of storing infinite information in a finite volume. Gisin thus proposed removing any physical meaning of real numbers, and he successfully showed that this implies that even classical mechanics ought to be interpreted indeterministically, in a similar fashion as advocated by Popper himself (Popper 1950). Subsequently, a new model was proposed in (Del Santo and Gisin 2019) which makes explicit use of propensities, which are taken to be tendencies of each digit of a physical quantity to take one of the possible but not yet determinate values. In this model, propensities are posited to be rational numbers and do not necessarily fulfil the formal requirements of probabilities. In this way, classical physics is modelled as an indeterministic (though empirically equivalent) theory, in which the values of physical quantities are not predetermined, but get actualized as time elapses, through a true random, objective process, i.e. a propensity.

In conclusion, Popper's contributions to physics are neither free of misconceptions nor of formal mistakes. Yet Popper's ability to always think out of the box, his stunning conceptual clarity and simplicity of explanation, together with his ability to engage in a multiplicity of subjects (and be able to interact with their respective practitioners) provided the ground for what we believe will be a long-lasting influence on the foundations of physics. Max Jammer (1991) commented that "Popper's wit, ingenuity, and independence of thought, [...] can undoubtedly have a stimulating effect on contemporary theoretizing in physics". Moreover, we cannot avoid looking in wonder at a scholar whose consistent vision spanned all disciplines, from politics and epistemology to physics and biology, a vision of hope in the intellectual freedom of humankind. At the same time, he was able to envision that nature itself does not rule out such freedom even for its most fundamental components, such as quantum particles: "The future is open. It is not predetermined and thus cannot be predicted –except by accident. The possibilities that lie in the future are infinite" (Popper 1994).

References

Bacciagaluppi, G., Valentini, A.: Quantum theory at the crossroads: reconsidering the 1927 solvay conference. Cambridge University Press, Cambridge (2009)

Ballentine, L.E.: Statistical interpretation of quantum mechanics. Rev. Mod. Phys. **42**(4), 358–381 (1970)

Ballentine, L.E.: Propensity, probability, and quantum theory. Found. Phys. **46**, 973–1005 (2016)

Birkoff, G., von Neumann, J.: The logic of quantum mechanics. Ann. Math. **37**(4), 823–843 (1936)

Bohm, D.: Quantum theory. Prentice-Hall, New York (1951)

Bohm, D.: A suggested interpretation of the quantum theory in terms of hidden variables—I & II. Phys. Rev. **85**(2), 166–179 and 180–193 (1952)

Broglie, L.: Nouvelles perspectives en microphysique. Albin Michel, Paris (1956)

Brush, S., Segal, A.: Making 20th century science: how theories became knowledge. Oxford University Press, New York (2015)

Bub, J.: Popper's propensity interpretation of probability and quantum mechanics. University of Minnesota Press, Minneapolis. Retrieved from the University of Minnesota Digital Conservancy. http://hdl.handle.net/11299/184671 (1972)

Del Santo, F.: Genesis of Karl Popper's EPR-like experiment and its resonance amongst the physics community in the 1980s. Stud. Hist. Philos. Mod. Phys. **62**, 56–70 (2018)

Del Santo, F.: Karl Popper's forgotten role in the quantum debate at the edge between philosophy and physics in 1950s and 1960s. Stud. Hist. Philos. Mod. Phys. **67**, 78–88 (2019)

Del Santo, F.: An unpublished debate brought to light: Karl Popper's enterprise against the logic of quantum mechanics. Stud. Hist. Philos. Mod. Phys. forthcoming (2020)

Del Santo, F., Cardelli, C.: Demolishing prejudices to get to the foundations: a criterion of demarcation for fundamentality. Foundations of Science (2019)

Del Santo, F., Gisin, N.: Physics without determinism: alternative interpretations of classical physics. Phys. Rev. A **100**(6) (2019)

Feyerabend, P.K.: On a recent critique of complementarity: part I. Philos. Sci. **35**, 309–333 (1968)

Freire, O. Jr. Popper, Probabilidade e Teoria Quântica. Episteme (Porto Alegre) **18**, 103–127. (2004)

Freire Jr., O.: The quantum dissidents: rebuilding the foundations of quantum mechanics (1950–1990). Springer, Berlin (2015)

Freire Jr., O.: David Bohm: a life dedicated to understanding the quantum world. Springer Nature, Switzerland (2019)

Garuccio, A., Popper, K.R., Vigier, J.-P.: An experiment to interpret E.P.R. action at a distance: the possible detection of real de Broglie waves. Epistemological Letters, **30**, 21 (1981a)

Garuccio, A., Popper, K.R., Vigier, J.-P. Possible direct physical detection of de Broglie waves. Physics Letters A, **86**(8), 397–400 (1981b)

Gisin, N.: Propensities in a non-deterministic physics. Synthese **89**, 287–297 (1991)

Gisin, N.: Indeterminism in physics, classical chaos and Bohmian mechanics. Are real numbers really real?. Erkenntnis (2019)

Gisin, N.: Real numbers are the hidden variables of classical mechanics. Quantum Stud.: Math. Found. **7**, 197–201 (2020)

Greenberger, D.M., et al.: Multiparticle interferometry and the superposition principle. Phys. Today **46**(8), 22–29 (1993)

Home, D., Whitaker, M.A.B.: Ensemble interpretations of quantum-mechanics—a modern perspective. Phys. Rep. Rev. Sect. Phys. Lett. **210**(4), 223–317 (1992)

Howard, D.: Popper and Bohr on realism in quantum mechanics. Quanta **1**(1), 33–57 (2012)

Humphreys, P.: Why propensities cannot be probabilities. Philos. Rev. **94**, 557–570 (1985)

Jammer, M.: The philosophy of quantum mechanics. Wiley, New York (1974)

Jammer, M.: Sir Karl Popper and his philosophy of physics. Found. Phys. **21**, 1357–1368 (1991)

Jarvie, I.C., Milford, K., Miller, D.: Karl Popper: a centenary assessment vol. 3: science. Ashgate Publishing Limited, Burlington (2006)

Kim, Y.-H., Shih, Y.: Experimental realization of Popper's experiment: violation of the uncertainty principle? Found. Phys. **29**(12), 1849–1861 (1999)

Kragh, H.: The most philosophically of all the sciences: Karl Popper and physical cosmology. Perspect. Sci. **21**, 3 (2013)

Maxwell, N.: Is the quantum world composed of propensitons? In: Suárez, M. (ed.) Probabilities Causes and Propensities in Physics. Springer, Dordrecht (2011)

Miller, D.: Critical rationalism: a restatement and defense. Open Court Publishing Company, Chicago, and La Salle (1994)

Popper, Karl R.: Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft. Julius Springer, Vienna (1934)

Popper, K.R.: Indeterminism in quantum physics and in classical physics. Part I. Br. J. Philos. Sci. **1**(2), 117–133 (1950)

Popper, K.R.: Indeterminism in quantum-mechanics and in classical physics. Br. J. Philos. Sci. **1**(2), 117–133 (1951).

Popper, K.R.: The propensity interpretation of the calculus of probability, and the quantum theory. In: Körner, S., Pryce, N.H.L. (eds.) Observation and Interpretation, Proceedings of the Ninth Symposium of the Colston Research Society (1957)

Popper, K.R.: The propensity interpretation of probability. Br. J. Philos. Sci. **10**(37), 25–42 (1959)

Popper, K.R.: Quantum mechanics without the observer. In: Bunge, Mario (ed.) Quantum Theory and Reality. Springer, Berlin (1967)

Popper, K.R.: Birkhoff and von Neumann's interpretation of quantum mechanics. Nature **219**, 682–685 (1968)

Popper, K.R.: Particle annihilation and the argument of Einstein, Podolsky and Rosen. In: Yourgrau, W., van der Merwe, A. (eds.) Perspectives in Quantum Theory. MIT Press, Cambridge (1971)

Popper, K.R.: Unended quest: an intellectual autobiography. Routledge, London (1976)

Popper, K.R.: The myth of the framework: in defence of science and rationality. Routledge, London and New York (1994)

Popper, K.R., Bartley III, W. (eds.): Vol. III of the Postscript to the Logic of Scientific Discovery: Quantum Theory and the Schism in Physics. Rowman and Littlefield, Hutchinson, Totowa, London (1982)

Qureshi, T.: Popper's experiment: a modern perspective. Quanta **1**(1), 19–32 (2012)

Redhead, M.: Popper and the quantum theory. R. Inst. Philos. Suppl. **39**, 163–176 (1995)

Selleri, F.: Quantum paradoxes and physical reality. Kluwer, Dordrecht (1990)

Shields, W.M.: A historical survey of Sir Karl Popper's contribution to quantum mechanics. Quanta **1**(1), 1–12. (2012)

Shih, Y., Kim, Y.-H.: Experimental realization of Popper's experiment: violation of the uncertainty principle? Fortschr. Phys. **48**(5–7), 463–471 (2000)

Suárez, M.: Propensities in quantum mechanics. In: Greenberger, D., Hentschel, K., Weinert, F. (eds.) Compendium of Quantum Physics. Springer, Berlin, Heidelberg (2009)

Tarozzi, G., van der Merwe, A. (eds.): Open questions in quantum physics. Reidel Publishing Co., Dordrecht (1985)

Flavio Del Santo is a doctoral candidate in theoretical physics at the University of Vienna and at the Institute of Quantum Optics and Quantum Information (IQOQI-Vienna) of the Austrian Academy of Sciences. His research focuses on fundamental issues of quantum physics—such as the measurement problem and the concept of indeterminism in quantum and classical physics—as well as the history and the philosophy of physics.

Olival Freire Jr. is Professor of History of Science at the Federal University of Bahia and Researcher at the National Council for Scientific and Technological Development (CNPq) in Brazil. He is the author of *The Quantum Dissidents—Rebuilding the Foundations of Quantum Mechanics 1950–1990*, Springer, 2015 and *David Bohm—A Life Dedicated to Understanding the Quantum World*, Springer, 2019. His main current interest is on the history of quantum physics, history of physics in Brazil, and the uses of history and philosophy in science teaching.

Popper's Experiment

Yanhua Shih

In 1934, one year before the 1935 paper of Einstein–Podolsky–Rosen (Einstein et al. (1935)), Popper published a thought experiment to probe the foundation of quantum theory according to his philosophy of realism (Popper 1934). Popper's original thought experiment is schematically shown in Fig. 1. A point source S, positronium as Popper suggested, is placed at the center of the experimental arrangement from which entangled pair of particle 1 and particle 2 are emitted in opposite directions along the respective positive and negative x-axes towards two screens A and B. There are slits on both screens parallel to the y-axis and the slits may be adjusted by varying their widths Δy. Beyond the slits on each side stand an array of Geiger counters for the joint measurement of the particle pairs as shown in the figure. The entangled pair could be emitted to any direction in 4π solid angles from the point source. However, if particle 1 is detected in a certain direction, particle 2 is then known to be in the opposite direction due to the momentum conservation of the entangled particle pair.

First, let us imagine the case in which slits A and B are both adjusted very narrowly. In this circumstance, particle 1 and particle 2 experience diffraction at slit-A and slit-B, respectively, and exhibit greater Δp_y for smaller Δy of the slits. There seems to be no disagreement in this situation between Copenhagen and Popper.

Next, suppose we keep slit-A very narrow and leave slit-B wide open. The main purpose of the narrow slit-A is to provide the precise knowledge of the position y of particle 1 and this subsequently determines the precise position of its twin (particle 2) on side B through quantum entanglement. Now, Popper asks, in the absence of the physical interaction with an actual slit, does particle 2 experience a greater uncertainty in Δp_y due to the precise knowledge of its position? Based on his beliefs of realism, Popper provides a straightforward prediction: *particle 2 must not experience a greater* Δp_y *unless a real physical narrow slit-B is applied.* However, if Popper's conjecture is correct, this would imply the product of Δy and Δp_y of particle 2 could be smaller than h ($\Delta y \, \Delta p_y < h$). This may pose a serious difficulty for Copenhagen and perhaps for many of us. On the other hand, if particle 2 going

---

Y. Shih (✉)
Department of Physics, University of Maryland, Baltimore, MD 21250, USA
e-mail: shih@umbc.edu

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_3

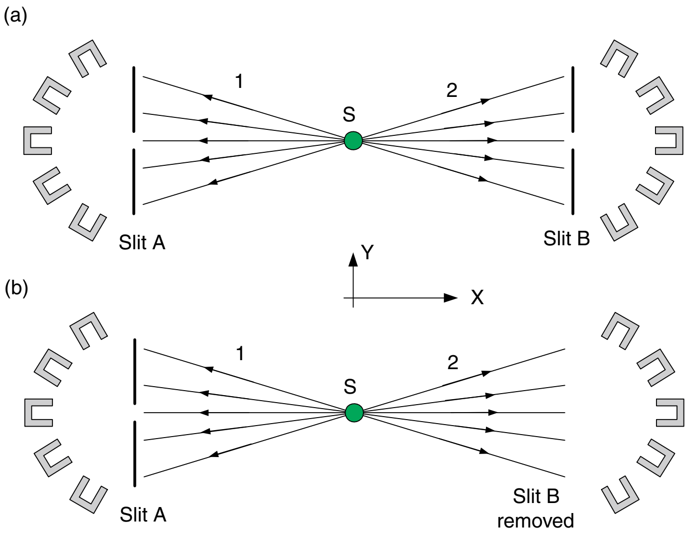


Fig. 1 Popper's thought experiment. An entangled pair of particles are emitted from a point source with momentum conservation. A narrow slit on screen A is placed in the path of particle 1 to provide the precise knowledge of its position on the y-axis and this also determines the precise y-position of its twin, particle 2, on screen B. **a** Slits A and B are both adjusted very narrowly. **b** Slit-A is kept very narrow and slit-B is left wide open

to the right does scatter like its twin, which has passed though slit-A, while slit-B is wide open, we are then confronted with an apparent *action-at-a-distance*!

The use of a *point source* in Popper's proposal has been criticized historically as a fundamental error Popper made. It is true that a point source can never produce a pair of entangled particles which preserves Einstein–Podolsky–Rosen (EPR) correlation (Einstein et al. (1935)) in momentum as Popper expected. However, notice that a *point source* is *not* a necessary requirement for Popper's experiment. What is required is a precise position-position EPR correlation: if the position of particle 1 is precisely known, the position of particle 2 is 100% determined. Ghost imaging is a perfect tool to achieve this.

Popper's Experiment One

Popper's thought experiment was realized by Kim and Shih in 1999 (Kim and Shih 1999) with the help of biphoton ghost imaging (Pittman et al. 1995). Figure 2 is a unfolded schematic diagram of their experiment, which is helpful for comparison with the original Popper's thought experiment.

This is a "unfolded" ghost imaging experimental setup (Pittman et al. 1995). An entangled signal-idler photon pair, generated from the nonlinear optical process of Spontaneous Parametric Down-Conversion (SPDC) of a BBO crystal, is used to image slit-A onto its distant ghost image plane of "screen" B. In the setup, s_o is chosen to be twice the focal length of the imaging lens LS, $s_o = 2f$. According to the Gaussian thin lens equation, an equal size "ghost" image of slit-A appears on the ghost image plane at $s_i = 2f$. The use of slit-A provides a precise knowledge of the position of photon 1 (signal) on the y-axis and also determines the precise y-position of its twin, photon 2 (idler), on screen B by means of the biphoton ghost imaging. The experimental condition specified in Popper's experiment is then achieved: when slit-A is adjusted to a certain narrow width and slit-B is wide open, slit-A provides precise knowledge about the position of photon 1 on the y-axis up to an accuracy Δy which equals the width of slit-A, and the corresponding ghost image of slit-A at screen B determines the precise position y of photon 2 to within the same accuracy Δy. Δp_y of photon 2 can be independently studied by measuring the width of its "diffraction pattern" at a certain distance from "screen" B. This is obtained by recording coincidences between detectors D_1 and D_2 while scanning detector D_2 along its y-axis, which is behind screen B at a certain distance.

Figure 3 is a conceptual diagram to connect the modified Popper's experiment of Kim and Shih (Kim and Shih 1999) with the biphoton ghost imaging experiment of

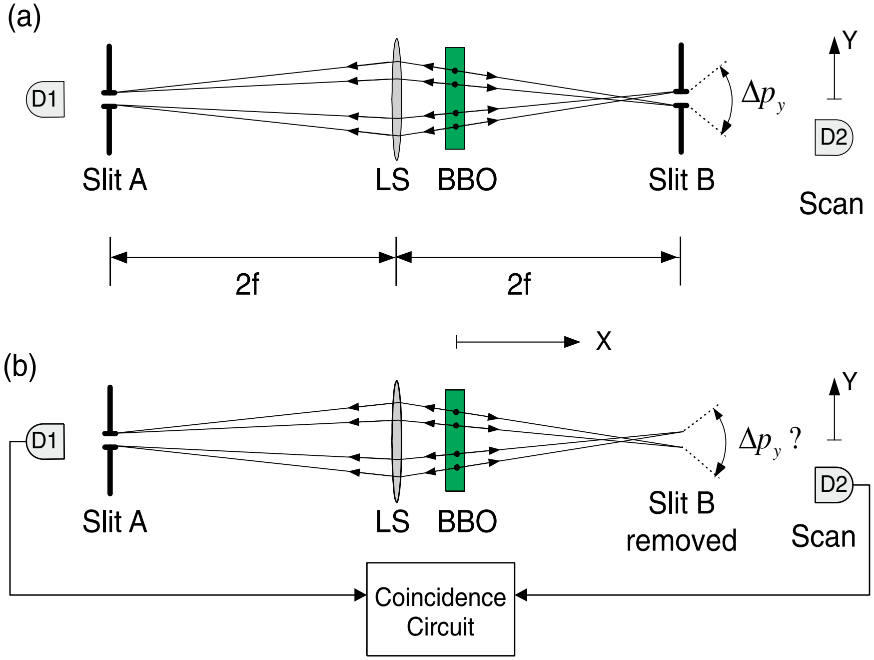


Fig. 2 Modified version of Popper's experiment. An entangled signal-idler photon pair is generated from the nonlinear optical process of Spontaneous Parametric Down-Conversion (SPDC) of a BBO crystal. A lens and a narrow slit-A are placed in the path of photon 1 to provide the precise knowledge of its position on the y-axis and also to determine the precise y-position of its twin, photon 2, on screen B by means of biphoton ghost imaging. Photon counting detectors D_1 and D_2 are used to scan in y-directions for joint detections. **a** Slits A and B are both adjusted very narrowly. **b** Slit-A is kept very narrow and slit-B is left wide open

Pittman *et al.* (Pittman et al. 1995). In this unfolded ghost imaging setup, we assume the entangled signal-idler photon pair holds a perfect EPR correlation in momentum with $\delta(\mathbf{k}_s + \mathbf{k}_i) \sim 0$, which can be easily achieved in a large transverse sized SPDC. In this experiment, we have chosen $s_o = s_i = 2f$. Thus, an equal size ghost image of slit-A is expected to appear on the ghost image plane of screen B.

The detailed experimental setup is shown in Fig. 4 with indications of the various distances. A CW Argon ion laser line of $\lambda_p = 351.1$ nm is used to pump a 3 mm long beta barium borate (BBO) crystal for type-II SPDC to generate an orthogonally polarized signal-idler photon pair. The laser beam is about 3 mm in diameter with a diffraction limited divergence. It is important not to focus the pump beam so that the phase-matching condition, $\mathbf{k}_s + \mathbf{k}_i = \mathbf{k}_p$, is well reinforced in the SPDC process, where $\mathbf{k}_j$ $(j = s, i, p)$ is the wavevectors of the signal (s), idler (i), and pump (p) respectively. The collinear signal-idler beams, with $\lambda_s = \lambda_i = 702.2nm = 2\lambda_p$ are separated from the pump beam by a fused quartz dispersion prism, and then split by a polarization beam splitter PBS. The signal beam (photon 1) passes through the converging lens LS with a 500 mm focal length and a 25 mm diameter. A 0.16 mm slit is placed at location A which is 1000 mm ($s_o = 2f$) behind the lens LS. A short focal length lens is used with D_1 for collecting all the signal photons that passe through slit-A. The point-like photon counting detector D_2 is located 500 mm behind "screen B". "Screen B" is the ghost image plane defined by the Gaussian thin equation. Slit-B, either adjusted as the same size as that of slit-A or opened completely, is placed to coincide with the ghost image of slit-A. The output pulses from the detectors are interfaced with a coincidence counter. During the measurements, the bucket detector

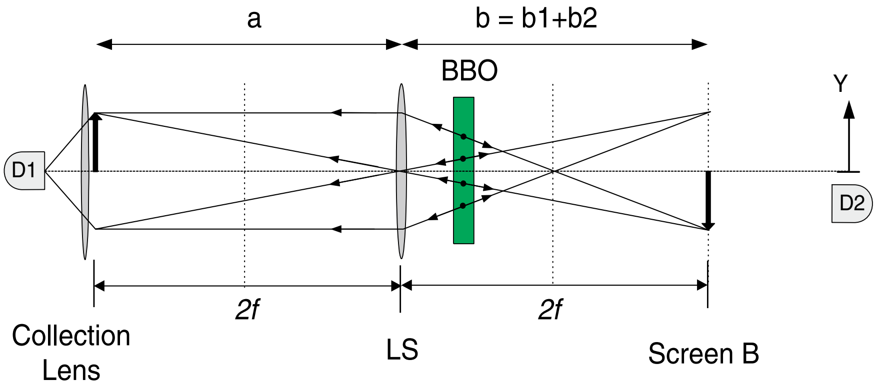


Fig. 3 An unfolded schematic of ghost imaging, which is helpful for understanding Kim and Shih's realization of Popper's experiment. We assume the entangled signal-idler photon pair holds a perfect momentum correlation $\delta(\mathbf{k}_s + \mathbf{k}_i) \sim 0$. The locations of the slit-A, the imaging lens LS, and the ghost image must be governed by the Gaussian thin lens equation. In this experiment, we have chosen $s_o = s_i = 2f$. Thus, the ghost image of slit-A is expected to be the same size as that of slit-A

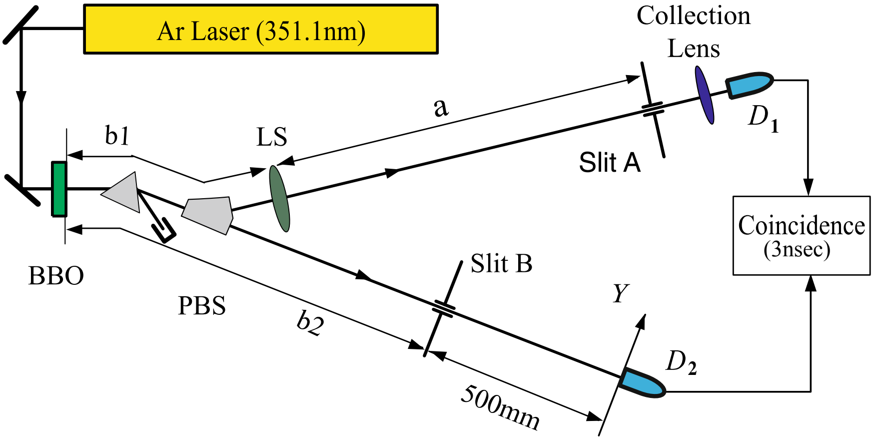


Fig. 4 Schematic of the experimental setup of Kim and Shih. The laser beam is about 3 mm in diameter. The "phase-matching condition" is well reinforced. Slit-A (0.16 mm) is placed 1000 mm $= 2f$ behind the converging lens, LS ($f = 500$ mm). The one-to-one ghost image (0.16 mm) of slit-A is located at B. The optical distance from LS in the signal beam taken as back through PBS to the SPDC crystal ($b_1 = 255$ mm) and then along the idler beam to "screen B" ($b_2 = 745$ mm) is 1000 mm $= 2f$ ($b = b_1 + b_2$)

D_1 is fixed behind slit-A while the point detector D_2 is scanned on the y-axis by a step motor.

Measurement 1: this measurement studied the case in which both slits A and B were adjusted to be 0.16 mm. The y-coordinate of D_1 was chosen to be 0 (center) while D_2 was allowed to scan along its y-axis. The circled dot data points in Fig. 5 show the *coincidence* counting rates against the y-coordinates of D_2. It is a typical single-slit diffraction pattern with $\Delta y \, \Delta p_y = h$. Nothing is special in this measurement except that we have learned the width of the diffraction pattern for the 0.16 mm slit and this represents the minimum uncertainty of Δp_y. We should emphasize at this point that the *single* detector counting rate of D_2 as a function of its position y is basically the same as that of the coincidence counts except for a higher counting rate.

Measurement 2: the same experimental conditions were maintained except that slit-B was left wide open. This measurement is a test of Popper's prediction. The y-coordinate of D_1 was chosen to be 0 (center) while D_2 was allowed to scan along its y-axis. Due to the entangled nature of the signal-idler photon pair and the use of coincidence counter, only those twins which have passed through slit-A and the "ghost image" of slit-A at screen B with an uncertainty of $\Delta y = 0.16$ mm (which is the same width as the real slit-B we have used in measurement 1) would contribute to the coincidence counts through the joint detection of D_1 and D_2. The diamond dot data points in Fig. 5 report the measured coincidence counting rates against the y coordinates of D_2. The measured width of the pattern is narrower than that of the diffraction pattern shown in measurement 1. It is also interesting to notice that the

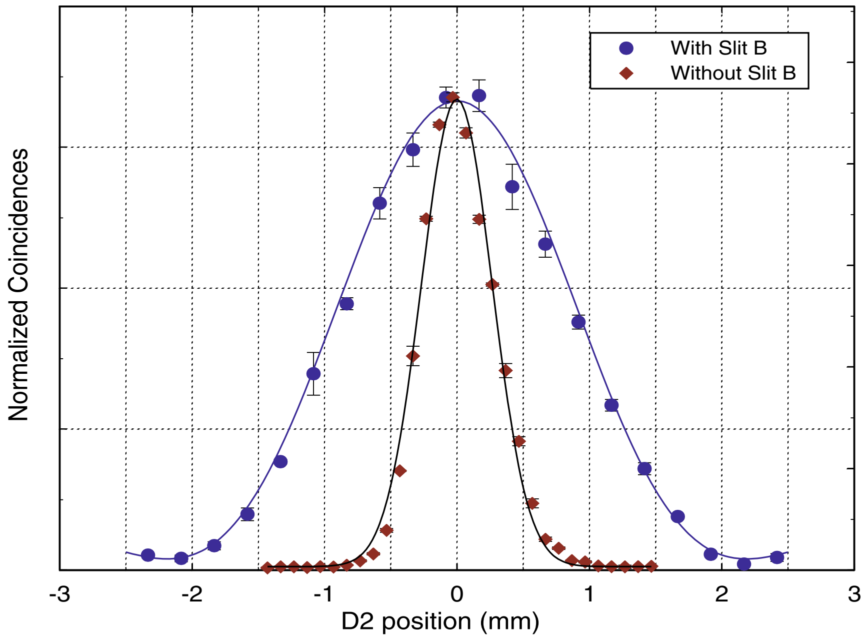


Fig. 5 The observed coincidence patterns. The y-coordinate of D_1 was chosen to be 0 (center) while D_2 was allowed to scan along its y-axis. Circled dot points: slit-A $= 0.16$ mm, slit-B $= 0.16$ mm. Diamond dot points: slit-A = 0.16 mm, slit-B wide open. The width of the sinc-function curve fitted by the circled dot points is a measure of the minimum Δp_y determined by a 0.16 mm slit. The fitting curve for the diamond dots is numerical result of Eq. (1), indicating a *blurred* ghost image of silt A

single detector counting rate of D_2 keeps constant in the entire scanning range, which is very different from that of measurement 1. The experimental data has provided a clear indication of $\Delta y \, \Delta p_y < h$ in the joint measurement of the entangled photon pairs.

Given that $\Delta y \, \Delta p_y < h$, is this a violation of the uncertainty principle? Does quantum mechanics agree with this peculiar experimental result? If quantum mechanics does provide a solution with $\Delta y \, \Delta p_y < h$ for photon 2, we would indeed be forced to face a paradox as EPR had pointed out in 1935 (Einstein et al. 1935).

Quantum mechanics does provide a solution that agrees with the experimental result. We now examine the experimental results from the view of quantum mechanics:

(1): When slits-A $= 0.16$ mm, slit-B $= 0.16$ mm.

This is the experimental condition for measurement one: slit-B is adjusted to be the same as slit A. There is no surprise for this measurement. The measurement simply provide us the knowledge of Δp of photon 2 after the diffraction coursed by slit-B of $\Delta y = 0.16$ mm. The experimental data shown in Fig. 5 agrees with the calculation. Notice that slit-B is about 745 mm far away from the 3 mm two-photon source, the angular size of the light source is roughly the same as $\lambda/\Delta y$, $\Delta\theta \sim \lambda/\Delta y$, where $\lambda = 702$ nm is the wavelength and $\Delta y = 0.16$ mm is the width of the slit. The calculated diffraction pattern is very close to that of the "far-field" Fraunhofer diffraction of a 0.16 mm single-slit.

(2): When slit-A $= 0.16$ mm, slits-B $\sim\infty$ (wide open).

Now we remove slit-B (wide open) from the ghost image plane. The quantum mechanical calculation of the transverse effective two-photon wavefunction and the second-order correlation is the same as that of the ghost image except the observation plane of D_2 is moved from the image plane a distance of 500 mm behind (Shih 2011). The two-photon image of slit A is located at a distance $s_i = 2f = 1000\,\text{mm}\,(b_1 + b_2)$ from the imaging lens, in this measurement D_2 is placed at $d = 1500$ mm from the imaging lens. The measured pattern is simply a "blurred" two-photon image of slit A. The "blurred" two-photon image can be calculated from Eq. (1)

$$\Psi(\boldsymbol{\rho}_o, \boldsymbol{\rho}_2) \propto \int_{lens} d\boldsymbol{\rho}_l\, G(|\,\boldsymbol{\rho}_2 - \boldsymbol{\rho}_l\,|, \frac{\omega}{cd})\, G(|\boldsymbol{\rho}_l|, \frac{\omega}{cf})\, G(|\,\boldsymbol{\rho}_l - \boldsymbol{\rho}_o\,|, \frac{\omega}{cs_o})$$

$$\propto \int_{lens} d\boldsymbol{\rho}_l\, G(|\,\boldsymbol{\rho}_l|, \frac{\omega}{c}[\frac{1}{s_o} + \frac{1}{d} - \frac{1}{f}])\, e^{-i\frac{\omega}{c}(\frac{\boldsymbol{\rho}_o}{s_o} + \frac{\boldsymbol{\rho}_i}{d})\cdot\boldsymbol{\rho}_l} \tag{1}$$

where d is the distance between the imaging lens and D_2. In this measurement, D_2 was placed 500 mm behind the image plane, i.e., $d = s_i + 500$ mm. The numerical calculated "blurred" image, which is narrower then that of the diffraction pattern of the 0.16 mm slit-B, agrees with the measured result of Fig. 5 within experimental error.

The measurement does show a result of $\Delta y\, \Delta p_y < h$. The measurement, however, has nothing to do with the uncertainty relation that governs the behavior of photon 2 (the idler). Popper was correct in the prediction of the outcomes of his experiment. Popper, on the other hand, made an error by applying the physics of two-particle system to the explanation of the behavior of an individual particle.

In Popper's experiment, the measurements are *joint detection* between two detectors applied to entangled states. Quantum mechanically, an entangled two-particle state only provides *the precise knowledge of the correlations of the pair*. The behavior of *photon 2* observed in the joint measurement is conditioned upon the measurement of its twin. A quantum must obey the uncertainty principle but the *conditional behavior* of a quantum in an entangled biparticle system is different in principle. We believe paradoxes are unavoidable if one insists the *conditional behavior* of a particle is the *behavior* of the particle. This is the central problem in the rationale behind Popper. $\Delta y\, \Delta p_y \geq h$ is not applicable to the conditional behavior of either *photon 1* or *photon 2* in the experiment of Popper.

The behavior of photon 2 being conditioned upon the measurement of photon 1 is well represented by the two-photon amplitudes. Each of the *straight lines* in Fig. 3 corresponds to a two-photon amplitude. Quantum mechanically, the superposition of these two-photon amplitudes are responsible for a “click-click” measurement of the entangled pair. A “click-click” joint measurement of the two-particle entangled state projects out certain two-particle amplitudes, and only these two-particle amplitudes are featured in the quantum formalism. In the above analysis we never consider photon 1 or photon 2 *individually*. Popper’s question about the momentum uncertainty of photon 2 is then inappropriate. The correct question to ask in these measurements should be: what is the uncertainty of Δp_y for the signal-idler *pair* which are localized within $\Delta y = 0.16\,\text{mm}$ at “screen” A with and without slit-B? This is indeed the central point for Popper’s experiment.

Once again, the demonstration of Popper’s experiment calls our attention to the important message: the physics of the entangled two-particle system must inherently be very different from that of individual particles.

Popper’s Experiment Two

In fact, nonlocal ghost imaging is not only the property of entangled photon pairs; it can also be realized in the joint measurement of two randomly created and randomly paired photons in thermal state. In 2005, ten years after the first ghost imaging demonstration of Pittman *et al.*, a near-field lensless ghost imaging experiment that uses pseudo-thermal radiation source, was demonstrated by Valencia et. al. (Valencia et al. 2005). This experiment opened a door for the realization of Popper’s thought experiment through the joint measurement of a random pair of photons in thermal state.

With the help of a novel joint detection scheme, namely the photon number fluctuation correlation (PNFC) circuit (Chen et al. 2013), which distinguishes the positive and negative photon number fluctuations measured by two single-photon counting detectors, and calculates the correlation between them, we were able to produce the ghost image of an object at a distance with 100% visibility. By modifying the Kim–Shih experiment of 1999 with a different light source and a lensless configuration, Peng *et al.* realized Popper’s thought experiment again in 2015 (Peng et al. 2015). Figure 6 is an unfolded schematic of their experiment, in which a large enough angular sized thermal source produces an equal-sized ghost image of slit-A at the plane $d_B = d_A$. The ghost image of slit-A can be verified by scanning the point-like photodetector D_B in the plane of slit-B. This ghost image provides the value of Δy through the correlation measurement. Again, the question of Popper is: Do we expect to observe a diffraction pattern that satisfies $\Delta p_y \Delta y > h$? To answer this question, we again make two measurements following Popper’s suggestion. Measurement (1) is illustrated in the upper part of Fig. 6. In this measurement, we place slit-B, which has the same width as that of slit-A, coincident with the 1:1 ghost image of slit-A and measure the diffraction pattern by scanning D_B along the y-axis in the far-field of the ghost image. In this measurement, we learn the value of Δp_y due to the diffraction of a real slit of Δy. Measurement (2) is illustrated in the lower part of Fig. 6. Here, we open slit-B completely, scanning D_B again along the same y-axis to measure the “diffraction” pattern of the 1:1 ghost image with the same width as slit-A. By comparing the observed pattern width in measurement (2) with that of measurement (1), we can examine Popper’s prediction.

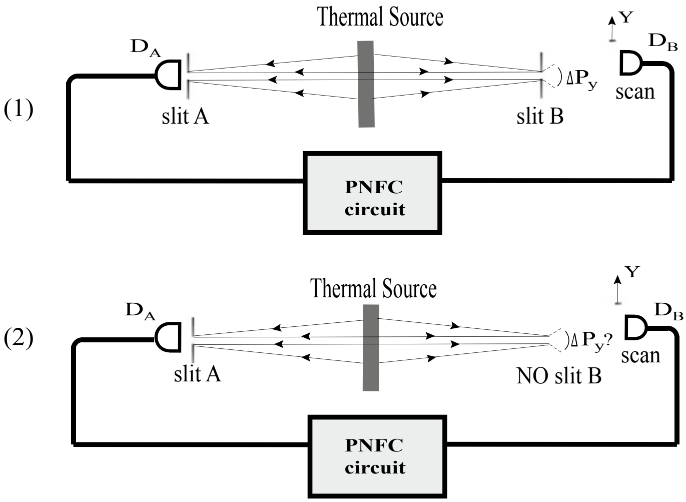


Fig. 6 Unfolded schematic of Popper's experiment with thermal light. The lensless ghost imaging setup with PNFC protocol produces an equal sized 100% visibility ghost image of slit-A at the position of slit-B. Detector D_B is scanning transversely in the y direction to measure the photon number fluctuation correlation with D_A when (1) Slit-A and slit-B are adjusted both very narrowly, and (2) Slit-A is kept very narrow and slit-B is left wide open

The experimental details are shown in Fig. 7. The light source is a standard pseudo-thermal source, consisting of a He-Ne laser beam and a rotating ground glass(GG). A 50/50 beamsplitter (BS) is used to split the pseudo-thermal light into two beams. One of the beams illuminates a single slit, slit-A, of width $D = 0.15$ mm located $d_A \sim 400$ mm from the source. A "bucket" photodetector D_A is placed right behind slit-A. An equal-sized ghost image of slit-A is then observable from the positive-negative photon number fluctuation correlation measurement between the "bucket" detector D_A and the transversely scanning point-like photodetector D_B, if D_B is scanned on the ghost image plane located at $d_B = d_A = 400$ mm. In this experiment, however, D_B is scanned on a plane that is located $d'_B \sim 900$ mm behind the ghost image plane, to measure the "diffraction" pattern of the ghost image. The output pulses from the two single-photon counting detectors are then sent to a PNFC circuit, which starts from two Pos-Neg identifiers follow two event-timers distinguish the "positive-fluctuation" Δn^+, from the "negative-fluctuation" Δn^-, measured by D_A and D_B, respectively, within each coincidence time window. The photon number fluctuation-correlations of D_A-D_B: $\Delta R_{AB} = \langle \Delta n_A \Delta n_B \rangle$ is calculated, accordingly and respectively, based on their measured positive-negative fluctuations.

The experiment was performed in two steps after confirming the 1:1 ghost image of slit-A. In measurement (1), we place slit-B ($D = 0.15$ mm) coincident with the ghost image and move D_B to a plane at $d'_B \sim 900$ mm to measure the diffraction pattern of slit-B. In measurement (2), we keep the same experimental condition as that of measurement (1), except slit-B is set wide open.

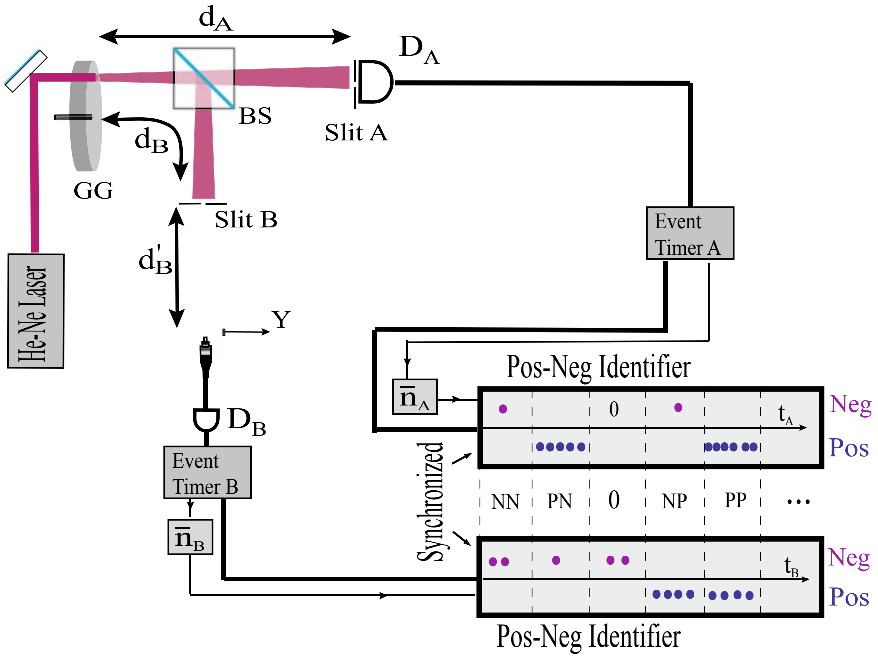


Fig. 7 Schematic of the experimental setup of Peng *et al.* A rotating ground glass (GG) is employed to produce pseudo-thermal light. BS is a 50/50 non-polarizing beam splitter. After BS, the transmitted beam passes through slit-A(0.15 mm) and collected by a "bucket" detector D_A which is put right after the slit. The reflected beam passes slit-B, which can be adjusted to be the same width as that of slit-A or wide open, and then reaches the scanning detector D_B. The distances from slit-A and slit-B to the source are the same($d_A = d_B = 400\,\text{mm}$). The distance from the scanning fiber tip of D_B to the plane of slit-B is $d'_B = 900\,\text{mm}$. A PNFC protocol is followed to evaluate the photon number fluctuation correlations from the coincidences between D_A and D_B

Figure 8 reports the experimental results. The circles show the normalized photon number fluctuation correlation from the PNFC protocol against the position of D_B along the y-axis for Popper's measurement (1). As expected, we observed a typical single-slit diffraction pattern giving us the uncertainty in momentum, Δp_y^{real}. The squares show the experimental observation from the PNFC for Popper's measurement (2), when slit-B is wide open. The measured curves agree well with our theoretical fittings. We found the width of the curve representing no physical slit is much narrower than that of the real diffraction pattern, which agrees with Popper's prediction.

We give a simply analysis by expressing the state of the randomly created and randomly paired photons in coherent state representation. We now calculate the joint photodetection counting rate of D_A and D_B which is proportional to the second-order coherence function $G^{(2)}_{AB}$:

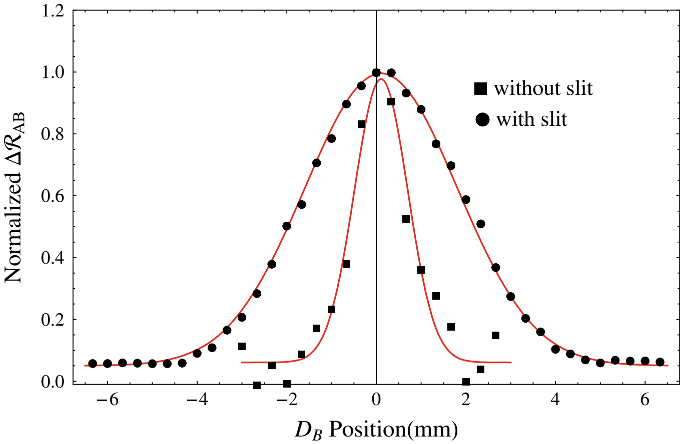


Fig. 8 The observed diffraction patterns. Circles: slit-A and slit-B are both adjusted for 0.15 mm. Squares: slit-A is 0.15 mm, slit-B is wide open. The width of the curve without the slit is almost three times narrower than that of the curve with slit, agreeing well with the theoretical predictions from Eqs. (7) and (9)

$$\begin{aligned} G_{AB}^{(2)} &= \Big\langle \langle \hat{E}^{(-)}(\boldsymbol{\rho}_A, z_A, t_A)\hat{E}^{(-)}(\boldsymbol{\rho}_B, z_B, t_B) \\ &\quad \times \hat{E}^{(+)}(\boldsymbol{\rho}_A, z_B, t_B)\hat{E}^{(+)}(\boldsymbol{\rho}_A, z_A, t_A)\rangle_{\text{QM}}\Big\rangle_{\text{Es}}, \end{aligned} \tag{2}$$

where $\hat{E}^{(+)}(\boldsymbol{\rho}_j, z_j, t_j)$ $(\hat{E}^{(-)}(\boldsymbol{\rho}_j, z_j, t_j))$ is the positive (negative) field operator at space-time coordinate $(\boldsymbol{\rho}_j, z_j, t_j)$, $j = A, B$, with $(\boldsymbol{\rho}_j, z_j, t_j)$ the transverse, longitudinal and time coordinates of the photodetection event of D_A or D_B. Note, in the Glauber–Scully theory (Scully 1997), the quantum expectation and classical ensemble average are evaluated separately, which allows us to examining the two-photon interference picture before ensemble averaging.

The field at each space-time point is the result of a superposition among a large number of subfields propagated from a large number of independent, randomly distributed and randomly radiating sub-sources of the entire thermal source,

$$\begin{aligned} \hat{E}^{(\pm)}(\boldsymbol{\rho}_j, z_j, t_j) &= \sum_m \hat{E}^{(\pm)}(\boldsymbol{\rho}_{0m}, z_{0m}, t_{0m}) g_m(\boldsymbol{\rho}_j, z_j, t_j) \\ &\equiv \sum_m \hat{E}_m^{(\pm)}(\boldsymbol{\rho}_j, z_j, t_j), \end{aligned} \tag{3}$$

where $\hat{E}^{(\pm)}(\boldsymbol{\rho}_{0m}, z_{0m}, t_{0m})$ is the mth subfield at the source coordinate $(\boldsymbol{\rho}_{0m}, z_{0m}, t_{0m})$, and $g_m(\boldsymbol{\rho}_j, z_j, t_j)$ is the optical transfer function that propagates the mth subfield from coordinate $(\boldsymbol{\rho}_{0m}, z_{0m}, t_{0m})$ to $(\boldsymbol{\rho}_j, z_j, t_j)$. We can write the field operators in terms of the annihilation and creation operators:

$$\hat{E}_m^{(+)}(\boldsymbol{\rho}_j, z_j, t_j) = C \int d\mathbf{k}\, \hat{a}_m(\mathbf{k})\, g_m(\mathbf{k}; \boldsymbol{\rho}_j, z_j, t_j), \tag{4}$$

where C is a normalization constant, $g_m(\mathbf{k}; \boldsymbol{\rho}_j, z_j, t_j)$, $j = A, B$, is the optical transfer function for mode $\mathbf{k}$ of the mth subfield propagated from the mth sub-source to the jth detector, and $\hat{a}_m(\mathbf{k})$ is the annihilation operator for the mode $\mathbf{k}$ of the mth subfield.

Substituting the field operators and the state, in the multi-mode coherent representation, into Eq. (2), we then write $G^{(2)}_{AB}$ in terms of the superposition of a large number of effective wavefunctions, or wavepackets:

$$\begin{aligned}
&G^{(2)}(\boldsymbol{\rho}_B, z_B, t_B; \boldsymbol{\rho}_B, z_B, t_B) \\
&= \Big\langle \sum_{m,n,p,q} \psi_m^*(\boldsymbol{\rho}_A, z_A, t_A)\psi_n^*(\boldsymbol{\rho}_B, z_B, t_B)\psi_p(\boldsymbol{\rho}_B, z_B, t_B)\psi_q(\boldsymbol{\rho}_A, z_A, t_A)\Big\rangle_{\text{Es}} \\
&= \Big\langle \sum_{m,n} \big|\psi_m(\boldsymbol{\rho}_A, z_A, t_A)\psi_n(\boldsymbol{\rho}_B, z_B, t_B) + \psi_n(\boldsymbol{\rho}_A, z_A, t_A)\psi_m(\boldsymbol{\rho}_B, z_B, t_B)\big|^2\Big\rangle_{\text{Es}} \\
&= \Big\langle \sum_{m} \big|\psi_m(\boldsymbol{\rho}_A, z_A, t_A)\big|^2 \sum_{n} \big|\psi_n(\boldsymbol{\rho}_B, z_B, t_B)\big|^2 \\
&\quad + \sum_{m\neq n}[\psi_m^*(\boldsymbol{\rho}_A, z_A, t_A)\psi_m(\boldsymbol{\rho}_B, z_B, t_B)\psi_n(\boldsymbol{\rho}_A, z_A, t_A)\psi_n^*(\boldsymbol{\rho}_B, z_B, t_B)]\Big\rangle_{\text{Es}} \\
&\equiv \langle n_A\rangle\langle n_B\rangle + \langle \Delta n_A \Delta n_B\rangle.
\end{aligned} \tag{5}$$

with

$$\psi_s(\boldsymbol{\rho}_j, z_j, t_j) = \int d\mathbf{k}\, \alpha_s(\mathbf{k}) e^{i\varphi_{0s}} g_s(\mathbf{k}; \boldsymbol{\rho}_j, z_j, t_j),$$

where $s = m, n, p, q,\ j = A, B$, and the phase factor $e^{i\varphi_{0s}}$ represents the random initial phase of the mth subfiled. In Eq. (5), we have completed the ensemble average in terms of the random phases of the subfields, i.e. φ_{0s}, and kept the nonzero terms only. Equation (5) indicates the second-order coherence function, is the result of a sum of a large number of subinterference patterns, each subpattern indicates an interference in which a random pair of wave packets interfering with the pair itself. For example, the mth and the nth wave packets have two different yet indistinguishable alternative ways to produce a joint photodetection event, or a coincidence count, at different space-time coordinates: (1) the mth wavepacket is annihilated at D_A and the nth wavepacket is annihilated at D_B; (2) the mth wavepacket is annihilated at D_B and the nth wavepacket is annihilated at D_A. In quantum mechanics, the joint detection probability of D_A and D_B is proportional to the normal square of the superposition of the above two probability amplitudes. We name this kind of superposition "nonlocal interference". The superposition of the two amplitudes for each random pair results in an interference pattern, and the addition of these large number of interference patterns yields the nontrivial correlation of the thermal light.

The cross interference term in Eq. (5) indicates the photon number fluctuation correlation $\langle \Delta n_A \Delta n_B \rangle$:

$$\begin{aligned} &\big\langle \Delta n_A(\boldsymbol{\rho}_A, z_A, t_A) \Delta n_B(\boldsymbol{\rho}_B, z_B, t_B) \big\rangle_{\text{Es}} \\ &= \Big\langle \sum_{m \neq n} \big[\psi_m^*(\boldsymbol{\rho}_A, z_A, t_A) \psi_n(\boldsymbol{\rho}_A, z_A, t_A)\big]\big[\psi_m(\boldsymbol{\rho}_B, z_B, t_B) \psi_n^*(\boldsymbol{\rho}_B, z_B, t_B)\big] \Big\rangle_{\text{Es}} \\ &\simeq \Big\langle \sum_m \psi_m^*(\boldsymbol{\rho}_A, z_A, t_A) \psi_m(\boldsymbol{\rho}_B, z_B, t_B) \sum_n \psi_n(\boldsymbol{\rho}_A, z_A, t_A) \psi_n^*(\boldsymbol{\rho}_B, z_B, t_B) \Big\rangle_{\text{Es}}. \end{aligned} \tag{6}$$

Measurement (1): slit-A = 0.15 mm, slits-B = 0.15 mm.

In measurement one, the optical transfer functions that propagate the fields from the source to D_A and D_B are

$$g_m(\boldsymbol{\kappa}, \omega; \boldsymbol{\rho}_A, z_A = d_A) = \frac{-i\omega e^{i(\omega/c)z_A}}{2\pi c d_A} \int d\boldsymbol{\rho}_s f(\boldsymbol{\rho}_s)\, e^{i\boldsymbol{\kappa}\cdot\boldsymbol{\rho}_s} G(|\boldsymbol{\rho}_s - \boldsymbol{\rho}_o|)_{[\omega/(cd_A)]},$$

and

$$\begin{aligned} &g_n(\boldsymbol{\kappa}, \omega; \boldsymbol{\rho}_B, z_B = d_B + d_B') \\ &= \frac{-\omega^2 e^{i(\omega/c)z_B}}{(2\pi c)^2 d_B d_B'} \int d\boldsymbol{\rho}_s \int d\boldsymbol{\rho}_i\; f(\boldsymbol{\rho}_s)\, e^{i\boldsymbol{\kappa}\cdot\boldsymbol{\rho}_s} G(|\boldsymbol{\rho}_s - \boldsymbol{\rho}_i|)_{[\omega/(cd_B)]} t(\boldsymbol{\rho}_i) G(|\boldsymbol{\rho}_i - \boldsymbol{\rho}_B|)_{[\omega/(cd_B')]}, \end{aligned}$$

where $\boldsymbol{\rho}_s$ is defined on the output plane of the source and $f(\boldsymbol{\rho}_s)$ denotes the aperture function of the source. We also assumed a perfect "bucket" detector D_A, which is placed at the object plane of slit-A ($\boldsymbol{\rho}_A = \boldsymbol{\rho}_o$), in the following calculation. $\boldsymbol{\rho}_i$ is defined on the ghost image plane, which is coincide with the plane of slit-B, and $\boldsymbol{\rho}_B$ is defined on the detection plane of D_B, $t(\boldsymbol{\rho}_i)$ is the aperture function of slit-B. The function $G(|\alpha|)_{[\beta]}$ is the Gaussian function $G(|\alpha|)_{[\beta]} = e^{i\frac{\beta}{2}|\alpha|^2}$. The measured fluctuation correlation can be calculated from Eq. (6)

$$\Delta R_{AB} = \int d\boldsymbol{\rho}_o\, |t(\boldsymbol{\rho}_o)|^2 \text{sinc}^2\Big[\frac{\omega_0 D \boldsymbol{\rho}_B}{2cd_B'}\Big] \equiv C' \times \text{sinc}^2\Big[\frac{\omega_0 D \boldsymbol{\rho}_B}{2cd_B'}\Big], \tag{7}$$

where $t(\boldsymbol{\rho}_o)$ is the aperture function of slit-A. The above calculation indicates a product between a constant C', which is from the integral on the "bucket" detector D_A, and a first order diffraction pattern of slit-B. With our experimental setup, the width of the diffraction pattern is estimated to be ∼4 mm, which agrees well with the experimental observation, as shown in Fig. 8.

Measurement (2): slit-A = 0.15 mm, slits-B ∼∞ (wide open).

In measurement two, slit-B wide open, the field at D_B becomes

$$g_n(\boldsymbol{\kappa}, \omega; \boldsymbol{\rho}_B, z_B) = \frac{-i\omega e^{i(\omega/c)z_B}}{2\pi c z_B} \int d\boldsymbol{\rho}_s f(\boldsymbol{\rho}_s) e^{i\boldsymbol{\kappa}\cdot\boldsymbol{\rho}_s} G(|\boldsymbol{\rho}_s - \boldsymbol{\rho}_B|)_{[\omega/(cz_B)]}.$$

We first check if a ghost image of slit-A is present when scanning D_B in the ghost image plane of $d_B = d_A$. The photon number fluctuation correlation is calculated to be:

$$\begin{aligned} \Delta R_{AB} &= \int d\boldsymbol{\rho}_o |t(\boldsymbol{\rho}_o)|^2 \mathrm{sinc}^2 \Big[\frac{\omega_0 a}{c d_A}|\boldsymbol{\rho}_o - \boldsymbol{\rho}_B|\Big] \\ &= |t(\boldsymbol{\rho}_o)|^2 \otimes \mathrm{sinc}^2 \Big[\frac{\omega_0 a}{c d_A}|\boldsymbol{\rho}_o - \boldsymbol{\rho}_B|\Big] \approx |t(\boldsymbol{\rho}_B)|^2. \end{aligned} \tag{8}$$

Note, we have placed D_A right behind slit-A and thus $\boldsymbol{\rho}_A = \boldsymbol{\rho}_o$. This suggests an equal-sized 100% visibility ghost image on the plane of $d_B = d_A$.

When we move D_B away from the ghost image plane to the far-field plane of $d_B + d_{B'}$, the photon number fluctuation correlation becomes:

$$\Delta R_{AB} = \int d\boldsymbol{\rho}_o |t(\boldsymbol{\rho}_o)|^2 \tilde{\mathcal{F}_s}^2\big(m\boldsymbol{\rho}_o - \boldsymbol{\rho}_B\big) = |t(\boldsymbol{\rho}_o)|^2 \otimes \tilde{\mathcal{F}_s}^2(m\boldsymbol{\rho}_o - \boldsymbol{\rho}_B), \tag{9}$$

where $\tilde{\mathcal{F}_s}$ is the of the defocused pupil function $\mathcal{F}_s = f(\boldsymbol{\rho}_s)e^{-i(\omega_0/2c\mu)\boldsymbol{\rho}_s^2}$ and μ, m are defined as $1/\mu = 1/d_A - 1/(d_B + d_{B'}), m = (d_B + d_{B'})/d_A$, respectively. The measured result of measurement (2) is thus a convolution between the aperture function of slit-A, $t(\boldsymbol{\rho}_o)$, and the correlation function $\tilde{\mathcal{F}_s}\big(m\boldsymbol{\rho}_o - \boldsymbol{\rho}_B\big)$, resulting in a "blurred" image of slit-A. With our experimental setup, the width of the "diffraction" pattern is estimated to be ~1.4 mm, which is almost three times narrower than the diffraction pattern of measurement (1) and agrees well with the experimental observation, as shown in Fig. 8. Compared with the Kim–Shih experimental result, we can see that although the number varies due to different experimental parameters, we have obtained a very similar result: the measured width of the "diffraction pattern" in measurement (2) is much narrower than that of the diffraction pattern in measurement (1).

The above analysis indicates that the experimental observations are reasonable from the viewpoint of the quantum coherence theory of light. The important physics we need to understand is to distinguish the first-order coherent effect and the second-order coherent effect, even if the measurement is for thermal light. In measurement (1), the fluctuation correlation is the result of first-order coherence. The joint measurement can be "factorized" into a product of two first-order diffraction patterns. After the integral of the "bucket" detector, which turns the diffraction pattern of slit-A into a constant, the joint measurement between D_A and D_B is a product between a constant and the standard first-order diffraction pattern of slit-B. There is no question the measured width of the diffraction pattern satisfies $\Delta p_y \Delta y \geq h$. In measurement (2) when slit-B is wide open or removed, the measurement can no longer be written as a product of single-photon detections but as a non-separable function, i.e., a convolution between the object aperture function and the photon number fluctuation correlation function of randomly paired photons, or the second-order coherence function of the thermal field. We thus consider the observation of $\Delta p_y \Delta y < h$ the result of the second-order coherence of thermal field which is caused from nonlocal two-photon interferences: a randomly paired photon interferes with the pair itself at a distance by means of a joint photodetection event between D_A and D_B. The result of nonlocal two-photon interference does not contradict the uncertainty principle that governs the behavior of single photons. Again, the observation of this experiment is not a violation of the uncertainty principle. The observation of $\Delta p_y \Delta y < h$ from thermal light, however, may reveals a concern about nonlocal interference as we have mentioned earlier.

References

Chen, H., Peng, T., Shih, Y.H.: 100% correlation of chaotic thermal light. Phys. Rev. A **88**, 023808 (2013)

Einstein, A., Podolsky, B., Rosen, N.: Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **47**, 777 (1935)

Kim, Y.H., Shih, Y.H.: Experimental realization of Popper's experiment: violation of the uncertainty principle? Found. Phys. **29**, 1849 (1999)

Peng, T., Simon, J., Chen, H., French, R., Shih, Y.H.: Popper's experiment with randomly paired photons in thermal state. Euro Phys. Lett. **109**, 14003 (2015)

Pittman, T.B., Shih, Y.H., Strekalov, D.V., Sergienko, A.V.: Optical imaging by means of two-photon entanglement. Phys. Rev. A **52**, R3429 (1995)

Popper, K.: Zur Kritik der Ungenauigkeitsrelationen. Naturwissenschaften **22**, 807 (1934)

Shih, Y.H.: An Introduction to Quantum Optics: Photon and Biphoton Physics. Taylor & Francis, CRC Press (2011)

Scully, M.O., Zubairy, M.S.: Quantum Optics. Cambridge University Press (1997)

Valencia A., Scarcelli G., D'Angelo M., Shih Y.H.: Two-photon imaging with thermal light. Phys. Rev. Lett. **94**, 063601(2005)

Yanhua Shih Professor of Physics, received his Ph.D. in 1987 from the Department of Physics, University of Maryland College Park, USA. He started the Quantum Optics Laboratory at the University of Maryland Baltimore County (UMBC) campus in the fall of 1989. His laboratory has been recognized as one of the leading groups in the field of quantum optics that attempts to probe the foundation of quantum theory. His pioneering research on multi-photon entanglement, multi-photon interferometry, quantum imaging, and optical tests of foundations of quantum theory has attracted a great deal of attention from the physics and engineering community. In thirty years of teaching and research, he published hundreds of experimental and theoretical works in leading refereed journals on the subject of quantum optics and has given hundreds of invited lectures and presentations at national and international professional conferences and workshops. Yanhua Shih received the Willis Lamb Medal in 2002 for his pioneer contributions to quantum optics and especially the study of coherence effects of multi-photon entangled states.

Karl Popper and Modern Cosmology: His Thoughts and Their Impact

Helge Kragh

> Physicists have this erotic obsession with Popperism and falsifiability criterion. (Keating 2017)

1 Introduction

Cosmology in the modern sense is sometimes said to have its origin in Einstein's static model of 1917 based on the general theory of relativity. During the next three decades the universe turned out to be expanding and the first ideas of an explosive beginning were proposed if not yet accepted. When Karl Popper died in 1994, the hot big-bang theory had been almost universally embraced for about thirty years, whereas the present picture of the universe known as the ΛCDM model was still in the future. Although Popper wrote very little about cosmology, he was interested in the subject and his sporadic writings on the subject are of interest at least from the perspective of history of science and ideas.

More importantly, Popper's philosophy of science and his falsifiability criterion in particular has played a surprisingly significant role in the discussions of post-World-War II cosmology. During the 1950s and 1960s Popper's ideas entered as an important element in the controversy over the then popular steady-state theory of the universe and they have continued to be part of the later development. In some of the more controversial areas of modern cosmology, as related to the multiverse hypothesis, inflation scenarios and the belief in dark matter, the ghost of Popper appears as if he were still alive. This chapter is divided in two parts, with the first one focusing on Popper's own views on cosmology and the second on how his philosophical ideas about science influenced and continues to influence discussions about the science of the universe.

H. Kragh (✉)
Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
e-mail: helge.kragh@nbi.ku.dk

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_4

2 Popper on Cosmology

2.1 Einstein's Influence on Popper

It is well known that Einstein's physics, and his general theory of relativity in particular, made a significant impact on Popper's philosophy of science (Kragh 2013). At some occasions Popper even stated that some of the central elements in his philosophy, such as falsifiability as a crucial criterion of science, were derived from Einstein's theories and his view of physics. What made a lasting impression on young Popper was Einstein's insistence that if just one of the predictions of his new theory of gravitation unambiguously disagreed with empirical evidence, the whole theory had to be abandoned. Popper stated that "Einstein's influence on my thinking has been immense" (Whitrow 1967, 23). On his side, Einstein recognized that Popper's philosophy as expounded in *Logik der Forschung* resonated with his own anti-induction view of science. Einstein read the book shortly after it appeared in late 1934, and in a letter to Popper of 15 June the following year he expressed how pleased he was with it.[1]

Popper was introduced to Einstein's general theory of relativity in around 1920, at a time when the theory aroused much public attention. As Popper recalled in his autobiography, he was impressed by Einstein's "marvellous idea of a new cosmology – a finite but unbounded universe" (Schilpp 1974, 28). However, neither Einstein's cosmological model of 1917 nor other aspects of modern cosmology appeared in *Logik der Forschung* (or in the later translation *The Logic of Scientific Discovery*). Indeed, in none of Popper's major works did he systematically address the scientific and philosophical problems of physical cosmology, a science which went through several revolutions during his lifetime. He did not ignore the field completely, though, and a close look at his books and papers reveals that he had interest as well as competence in questions of cosmology. Still, compared to his writings on other topics of theoretical physics, such as thermodynamics, statistical physics and quantum mechanics, his scattered remarks on scientific cosmology are few and easily overlooked.

Popper (1994, 59; 1959, 15) called cosmology "the most philosophically important of all the sciences" and stated that "all science is cosmology," but with these expressions he did not refer to the scientific version of cosmology cultivated by astronomers and physicists during the twentieth century. He used the term "cosmology" in its older and much broader sense of understanding the world, a sense which was also used by Stephen Toulmin (1982) and other philosophers contemporaneous with Popper. In his essays on the history of cosmological thought Popper characteristically disregarded developments in the twentieth century.

---

[1] Popper responded in a long letter of 18 July 1935. For the Einstein-Popper correspondence, see https://www.aau.at/universitaetsbibliothek-klagenfurt/sondersammlungen/kostbarkeiten-aus-der-bibliothek/einstein_popper/.

One example is a BBC broadcast he gave in 1954 on the occasion of the 150th anniversary of the death of Immanuel Kant and in which he highlighted the great philosopher's pioneering treatise *Allgemeine Naturgeschichte und Theorie des Himmels* from 1755. Popper found inspiration in Kant's discussion of whether the universe is finite or infinite in either a spatial or a temporary sense. As he pointed out, Einstein's model of "a world which is both finite and without limits" solved half of the problem, but "as far as time is concerned no equally promising solution of Kant's difficulties has been offered up to now" (Popper 1963, 178). Most likely the brief remarks reflected the ongoing controversy between relativistic evolution models and the new steady-state theory. They possibly indicate that at the time Popper favoured a closed universe but not a finite-age universe of the big-bang type. He did not find the latter theory to be "promising."

2.2 Does the Universe Expand?

Einstein's cosmological model of 1917 was static, but in 1930, when Popper was 28 years old, the revolutionary idea of the expanding universe entered the cosmological scene. Expanding models were theoretically justified as solutions to the cosmological field equations and empirically by the linear Hubble law relating the redshifts of galaxies to their distances. Although a majority of astronomers soon accepted the expanding universe, throughout the 1930s a significant minority argued that the redshifts might be explained on the basis of a static world (Kragh 2019). Popper was for a time interested in the latter kind of theories, such as he recalled in a letter to the author of 10 June 1994[2]:

> I had many ideas within and outside *Relativistic* cosmology, especially also about non-Hubble (= non-Doppler) explanations of the redshift: cp. my paper in 1940 in *Nature*, 145, pp. 69. … Later, I somewhere published the remark that sunlight suffers a redshift when the sun stands low, because of collision with particles, and that the same must happen in "empty" space, so *there must be a redshift increasing with distance.*

In the paper from 1940, written while he stayed in Christchurch, New Zealand, Popper (1940) responded to a recent suggestion that the redshifts were due to the speed of light decreasing in time, and he combined this hypothesis with the cosmological system proposed by the British physicist Edward A. Milne. He had obviously studied Milne's theory, which rested on the conventionalist hypothesis that the description of nature depends on the chosen scale of time. According to Milne, what appeared to be a universe expanding from a singularity might be transformed into a static world by using a logarithmic time scale. Inspired by Milne's theory, Popper argued that the non-expanding hypotheses were logically and empirically equivalent

---

[2] See Kragh (2012). The letter is deposited at the Karl Popper Collection, Klagenfurt University Library. I have not found Popper's remark of a "tired light" hypothesis for the cosmic redshift and suspect that it was never published.

to the standard expansion theory. They "do not describe alternative *facts*, but the same facts in alternative *languages*." Popper elaborated:

> To ask whether "in reality" the universe expands, or *c* [velocity of light] decreases, or the frequencies speed up, is not more legitimate than, when prices of goods fall throughout the economic system, to ask whether "in reality" the value of money has increased or the value of the goods has decreased.

It is hard to avoid the conclusion that, as far as the cosmic expansion is concerned, Popper's attitude was decidedly anti-realistic.

Despite his conventionalist approach Popper considered Milne's alternative to be a mathematically simpler and therefore more attractive explanation of the redshifts. Moreover, he used Milne's ideas to derive an expression for the time-dependence of the apparent luminosities of galaxies. Inspired by a recent paper of Erwin Schrödinger (1939) he essentially dealt with the so-called surface brightness test as a means of distinguishing between expanding and non-expanding models. As shown by the 1940 paper, Popper was at the time acquainted with and interested in cosmological research. However, the paper made no impact at all and apparently also Popper found it to be unimportant. In his autobiography, he chose not to mention it.

A further example of Popper's competence and on-and-off interest in cosmological theories is provided by his analysis of Kurt Gödel's time-symmetric, rotating and stationary world model (Schilpp 1974, 103 and 172). In 1950, he discussed Gödel's model with Einstein, who rejected it as an unphysical toy model. Although Popper apparently agreed and emphasized the reality of time in contrast to Gödel's idealistic view, he also found the model to be instructive as a philosophical lesson regarding topics such as global determinism and realism. Unfortunately, Popper's discussions with Einstein and Gödel in Princeton were not recorded.

2.3 Finite-Age Cosmological Models

According to most relativistic models, whether of the big-bang type or not, the universe could be ascribed a finite age. On the other hand, the rival steady-state universe was infinitely old. During the 1950s the question of the age of the universe was hotly debated, with Popper and several other philosophers taking an interest in it. In 1953 Popper sat on a prize committee established by the new *British Journal for the Philosophy of Science* to judge essays on the epistemic and scientific status of a finite-age universe (Kragh 2013). The first prize of the competition was shared by the American philosopher Michael Scriven and the British physical chemist John T. Davies. According to Scriven, present science was powerless to decide whether the age of the universe is finite or infinite. I do not know if Popper agreed or not, but the following year he expressed a somewhat similar view (see Sect. 2.1).

The 1953 competition is of relevance also because it illustrates the growing recognition in scientific circles of Popper's philosophy of science. Thus, Davies' essay referred explicitly to Popper's *Logik der Forschung* and repeated its message of falsifiability as a demarcation criterion for science. “As Popper has emphasized,” Davies (1954) wrote, “the criterion of a scientific theory is that it must be possible for an observational check to be devised … by which it might be disproved.”

The theme of the age of the universe turned up many years later in a brief exchange of arguments between Popper and his friend Gerald Whitrow, an astronomer who was also an expert on the philosophy of time. In 1978, at a time when the big-bang theory enjoyed almost universal acceptance, Whitrow offered logical reasons why an infinitely old universe is impossible. Although Popper (1978) did not defend an eternal universe, he thought it was possible and that Whitrow’s reasoning was unconvincing.[3] Popper’s conclusion reminds one of his paper of 1940, in so far that he appealed to Milne’s two time scales. Since a universe without a beginning could be transformed into one with a beginning, Popper doubted if there was any “ontological difference” between the two notions of cosmic duration. The Whitrow-Popper debate was philosophical and not scientific, and—somewhat strangely—none of the discussants referred to the strong empirical evidence for a universe born in a big bang.

Finally, Popper’s sure grasp of relativistic cosmology is substantiated in some of his works on time, entropy and irreversibility. In 1967 he discussed within this framework a semi-cyclic closed universe such as proposed by Einstein in 1931. More generally he dealt with the relationship between the average density of energy in the universe and the space curvature. Since this work is summarized and explicated in Kragh (2013) I shall not comment further on it except mentioning that Popper (1967) probably misinterpreted Einstein’s ideas.

2.4 Against the Big Bang

The now standard hot big-bang theory of the universe has roots in the early theories of Alexander Friedmann and Georges Lemaître but was only developed into a quantitative model by George Gamow and collaborators in the late 1940s (Kragh 1996). However, the majority of astronomers and physicists disregarded the big bang, which was only widely accepted in 1965 with the interpretation of the cosmic microwave background as a fossil of the early universe. Ten years later the big-bang theory had achieved an almost paradigmatic status in cosmology. Popper witnessed the dramatic development without ever being convinced that the universe had started, several billion years ago, in a dense inferno of high-energy radiation and nuclear particles.

---

[3] In a letter to Whitrow of 3 May 1977, Popper wrote: “I personally am happier with time without a beginning; but I do not argue for it … but against arguing a priori for one or the other.” Cited with the permission of Karl Popper Collection, Klagenfurt University Library, Austria (box 361, file 21). https://www.aau.at/en/university-library-klagenfurt/karl-popper-collection/.

At the end of his life Popper believed that the celebrated big-bang theory was nothing but an unscientific speculation, an untestable myth. In his letter of 1994 (see note 2), he referred to "the inexplicability of a beginning of time" and further wrote[4]:

> I was a (student) member at the [Vienna] department of theoretical physics when … Friedmann suggested that Hubble's suggestion could be explained by a simplified Einstein cosmology … [The] big bang theory became rapidly more and more complicated. And my present view is that the number of auxiliary hypotheses is simply intolerable: according to my theory of science, *this is not science*. … And not only is it not stressed by the upholders of the theory that it is all speculation without tests, but it is presented as if the theory were a proven *fact*. This is horrid; impermissible; against scientific ethics. … I once *was* an enthusiastic admirer of (Friedmann's) Big Bang. I am *now* a disgusted opponent. By contrast, Einstein's General Relativity is a marvelous theory.

In his writings and public lectures, Popper rarely addressed the new cosmological theories, but on a few occasions he indicated his sceptical view not only of the big-bang theory but also of physical cosmology generally. In an address delivered in Vienna in 1968 he expressed doubts about the very possibility of a science of the universe in its entirety. As late as 1992, in a letter to Milne's daughter and biographer, he characterized cosmology as "a largely speculative subject" (Weston Smith 2013, 156).

Popper wondered in his autobiography about "the infinitely improbable" success of physical cosmology (Schilpp 1974, 1027). The problem was the invalid inductive method on which cosmology must rely. After all, we only have empirical access to a small part of the universe, so how can we gain knowledge of the entire universe without relying on inductive generalization? "Modern cosmology," Popper wrote, "teaches us that to generalize from observations taken, for the most part, in our incredibly idiosyncratic region of the universe would almost always be quite invalid."

At some occasions Popper expressed his dislike of the cosmological principle (CP) on which standard cosmology relied, namely the assumption that on a sufficiently large scale the universe is homogeneous and isotropic. This principle was part of Einstein's model of 1917, and it was formulated as a formal principle by Milne some twenty years later. According to the CP, our region of the universe is not "incredibly idiosyncratic" but on the contrary representative for the universe as a whole, for which reason it is sometimes called a "principle of mediocrity." However, Popper thought it was a postulate beyond observational tests, an unwarranted generalization. A few prominent astronomers, such as Gérard de Vaucouleurs and Victor Ambartsumian, shared his dissatisfaction with the CP, but most considered it (and still consider it) to be justified by observations as well as theory.

Popper's attitude to modern cosmology may perhaps best be described as sceptical and agnostic. He found it a most interesting area of research but one which was not and might perhaps never become truly scientific. In a lecture of 1972, he summarized his view as follows: "Both cosmology and cosmogony, though immensely fascinating parts of physics, and though they are becoming better testable, are still almost borderline cases of physical science and hardly yet mature enough to serve as the bases of

---

[4] Popper's memory failed him, as Friedmann's prediction of an expanding universe dates from 1922 and Hubble's observations only from 1929. Also Lemaître's prediction was prior to Hubble's work.

the reduction of chemistry to physics."[5] Ten years later, in another lecture, Popper (1994, 60) reviewed the recent history of cosmology, this time without referring to the victorious big-bang theory. Instead he concluded that "we seem to be almost as helpless in the field of cosmology in the face of some of these revolutionary results as we are in politics when faced with the task of making peace."

Although Popper followed the development of cosmological research, he did it at a distance and without sharing the enthusiasm of many astronomers and physicists. To them, and to the public at large, the discovery of the cosmic microwave background was a revelation and solid proof of the big-bang theory. Remarkably, Popper ignored the background radiation and also, by and large, other evidence in favour of the big bang. In one of his few references to the hot big-bang theory he mentioned the claim (as he called it) that "most of the helium … was produced within the very first minute of the existence of the expanding universe." Popper (1982, 143) added that, "The precariousness of the scientific status of this speculation … need not be stressed." Popper's dissatisfaction with mainstream cosmology did not turn him into an advocate of some alternative theory of the universe. He knew about the steady-state theory but realized that latest by 1970 it was ruled out observationally. Still, from a methodological point of view he preferred the steady-state theory, or some modification of it, over the big-bang theory because the first had a much higher degree of falsifiability.

3 Cosmologists on Popper

3.1 The Steady-State Controversy: Bondi and Popper

The steady-state theory appeared in 1948 as a radical alternative to big-bang cosmology and other evolutionary models based on general relativity (Kragh 1996). In the version introduced by Hermann Bondi and Thomas Gold, the theory rested on the basic assumption that the large-scale features of the universe have always looked and will always look the same—what they called the "perfect cosmological principle" or PCP. Whether in this version or the one proposed by Fred Hoyle, the constant-density universe expanded exponentially and therefore required continual creation of matter, if of the miniscule rate 10^{-43} g/s/cm^3. The new theory resulted in a number of sharp predictions, which made it vulnerable to refutation, whereas the class of relativistic theories did not allow predictions that could be tested by some *experimentum crucis*. It was in this context that Popper's philosophy came to play an important role in the cosmological controversy which lasted from 1948 to about 1965.

---

[5] Popper (1982, 143), where he discussed reductionism generally, including the reduction of chemistry to physics. He also referred to the possibility of reducing physics to cosmogony, and in this context he mentioned the unorthodox theories of Paul Dirac and Pascual Jordan based on the hypothesis of a decreasing constant of gravitation (Kragh 2019).

Bondi was particularly fascinated by Popper's philosophy of science which turned up repeatedly in his methodological arguments for the steady-state theory. Like Popper an Austrian immigrant, he was thoroughly acquainted with Popper's views years before they made a wide impact with the publication of *The Logic of Scientific Discovery*. Together with the mathematical physicist Clive Kilmister, Bondi wrote in 1959 a glowing review of the book in *British Journal for the Philosophy of Science* (vol. 10, pp. 55–57). According to the two physicists, "Popper speaks as a working scientist to the working scientist in a language that time and again comes straight out of one's heart."

Popperian themes also appeared prominently in the Joule Memorial Lecture which Bondi gave in Manchester in 1958. On this occasion he summarized Popper's method as follows: "The theory, if it is to be a scientifically useful theory, must positively stick out its neck in order to run the risk of being disproved. … Only a vulnerable theory that suggests ways by which it can be disproved is fertile and of scientific use" (Bondi 1958, 60). At the end of his lecture Bondi referred to the cosmological controversy and the possibility of obtaining scientific knowledge of the universe as a whole. The only way to discriminate between rival conceptions of the universe was by means of critical tests in the sense of Popper, and in this regard Bondi was confident that the steady-state theory was superior. "Our theories are not cranky invulnerable speculations," he said, "they are proper scientific theories which suggest experiments by which they may be shot down."

Latest by 1960, Bondi had become an enthusiastic advocate of Popper's philosophy, which he praised in both a general sense and in relation to the uncertain state of cosmological theory. To give just one more example, in a BBC broadcast on modern cosmology from 1959 he characterized Popper's work as "by far the most successful analysis of scientific method" (Bondi et al. 1960, 12). Later in life, after the steady-state theory had become obsolete, Bondi continued to spread the gospel of Popper's philosophy of science and his world view generally. In an obituary in *Nature* (vol. 371, p. 478) he described his first meeting with Popper's thoughts as if he were hit by "a flash of brilliant light." Bondi and Popper met frequently and among the topics they discussed was the situation in cosmology. Popper was thus well informed about the steady-state theory and its fight against relativistic evolution theories. However, as he recalled in his letter of 1994, he did not read the scientific papers on the subject. Nor did he intervene in the philosophical debate over cosmology such as did several other philosophers, among them Adolf Grünbaum, Norwood Russell Hanson and Milton Munitz. Popper preferred to consider the controversy from the side-line.

Although Popper was to some extent sympathetic to the steady-state theory, he rejected the PCP, which according to Bondi served as the theory's very foundation. Bondi insisted that the PCP was liable to observational disproof and in this respect agreed with Popper's criterion of science. "This possibility of a clear-cut disproof establishes the scientific status of the P.C.P.," Bondi (1966, 396) wrote, but Popper nevertheless found the principle to be suspect. He thought it was a deplorable example of "making our *lack* of knowledge a principle of *knowing something*," as he phrased it in his letter of 1994. While Popper disliked the PCP as well as the ordinary CP limited to the spatial dimensions, he liked the element of matter creation in steady-state cosmology and thought that this should be the basic idea of the Bondi-Gold theory. A theory of this kind—a steady-state universe with matter creation but without the PCP—was proposed by Reginald Kapp, an engineer and writer on cosmology with whom Popper was in contact in the late 1950s (Kragh 1996, 196–197). Kapp informed Popper about his ideas and provided him with a copy of his book *Towards a Unified Cosmology* published in 1960, but it is unknown how Popper responded to Kapp's rather amateurish cosmological theory.

Popper's somewhat sceptical view of the scientific status of cosmology disagreed with the one expounded by Bondi. As mentioned, Popper was doubtful if the universe could be understood scientifically in the same way that limited physical systems could be understood. Bondi, on the other hand, held that existing physical cosmology was truly scientific, or at least on its way to become so, and he justified his epistemic optimism in the falsification criterion of science.[6] He tended to apply Popper's prescriptions in a rather strict way, as if comparison of theory and observations would always reveal whether the theory should survive or not. Bondi did not fully appreciate the finer details of Popper's theory, as when Popper (1959, 50) distanced himself from the concept of conclusive disproof and pointed out that falsifiability should be understood as a vague and not an absolute criterion (see Sect. 3.4).

Whitrow was critical to the steady-state theory without dismissing it, and yet he valued Popper's criterion of falsifiability as an important method in cosmology and science generally. The same was the case with some of the proponents of an evolving universe governed by general relativity, who did not consider Popperian standards to apply only to the steady-state theory. On the contrary, they maintained that relativistic models were and had to be empirically refutable, if admittedly not as easily as the steady-state model. Because Popperian standards of science were broadly accepted by both parties in the controversy, they were rarely an issue of dispute in the scientific literature. Popper's view of science was of course criticized by other philosophers, but astronomers and physicists were generally happy with it or they were just indifferent.

3.2 Falsifiability in Modern Cosmology

Although the Bondi-Gold version of steady-state theory was abandoned shortly after the discovery of the microwave background, Hoyle never admitted defeat. Together with Jayant Narlikar and a few other collaborators he produced a new series of steady-state theories which differed radically from the earlier PCP-inspired theories. Their efforts culminated with the proposal in the late 1990s of what they called

---

[6] This was the theme discussed by Whitrow and Bondi (1954), where Whitrow argued that physical cosmology would remain a borderland subject between science and philosophy. Both discussants referred approvingly to Popper's philosophy, Whitrow explicitly and Bondi implicitly. Popper's own view on cosmology seems closer to Whitrow's than to Bondi's.

the quasi-steady-state cosmology or QSSC, an eternally oscillating universe with periodic creations of matter (Hoyle et al. 2000). This theory claimed to account for all observational data, including the microwave background, and also to result in new testable predictions different from those of the standard big-bang theory. Hoyle was not a philosophical mind and contrary to Bondi he showed no interest in Popper's philosophical views. Nonetheless, he and other supporters of the QSSC research program were keen to point out that their theory was falsifiable and de facto in agreement with Popper's methodology.

In the third edition of his widely read textbook *Introduction to Cosmology*, Narlikar (2002, 497) raised the question of which tests can be performed that would disprove the QSSC alternative. "This question is in the spirit of Karl Popper's view of a scientific theory, namely that it should be disprovable," he stated. "If the theory seeks survival by adding an extra parametric dimension, then that is against the spirit of this question." As to the QSSC alternative, if non-baryonic candidates for dark matter were found, QSSC would be disproved, and the same would be the case if it were found that the universe had passed through epochs of very high redshifts ($z > 30$). The QSSC program turned out to be unsuccessful and after about 2010 it practically ceased.

QSSC is just one example of many illustrating that elements of Popper's philosophy are highly visible in modern astronomy and cosmology (Sovacool 2005; Kragh 2013). Consider what it probably the most best-selling book on cosmology ever, Stephen Hawking's *A Brief History of Science*. In the introductory chapter Hawking informed the readers about the nature of physical theory, which he essentially did by paraphrasing Popper: "As philosopher of science Karl Popper has emphasized, a good theory is characterized by the fact that it makes a number of predictions that could in principle be disproved or falsified by observation" (Hawking 1989, 11). Hawking further pointed out that although agreement with observations will increase our confidence in a theory, it will not prove it since other theories may result in the same predictions. On the other hand, "if ever a new observation is found to disagree, we have to abandon or modify the theory."

Although one of the fathers of the widely accepted inflation theory, the American physicist Paul Steinhardt came to the conclusion that cosmic inflation is unnecessary and of no scientific value. Among his and others' complaints are that inflation theory exists in numerous versions, and that inflation fails to result in unique predictions of such a kind that they can disprove the theory. Of course, the many supporters of inflation vehemently deny that this is the case or that inflation is somehow unscientific (see the dispute in the February and May 2017 issues of *Scientific American*). Together with Neil Turok, Steinhardt developed in the early years of the new millennium a cyclic model of the universe as an alternative to inflation big-bang cosmology. Without mentioning Popper by name, they emphasized that the new cyclic theory agreed with Popperian standards of falsifiability such as making predictions that, if they were proven wrong, would disprove the theory. In particular, whereas most inflation models predict an imprint in the density perturbations of the microwave background due to primordial gravitational waves, no such imprint should exist according to the cyclic theory. Steinhardt (2004, 469) considered this a crucial test à la Popper, as the detection would "*definitely* rule out the cyclic model."

Among the many exotic ideas of modern cosmology, the anthropic principle is possibly the most enduringly controversial. When Brandon Carter introduced the principle in 1973, he realized that it violated some of the cherished methods of physics, including Popper's emphasis on falsifiability and sharp predictions. However, it did not worry Carter too much, for he saw no reason why falsifiability should be rated higher than verifiability. He argued that if a consequence of a theory is confirmed and thus turned into a fact, it can no longer be refutable and yet it undoubtedly increases the credibility of the theory. Carter (1993, 51) was highly critical to the popular version of Popper's philosophy of science:

> By unfairly conveying the impression that confirmation is valueless unless absolute, without also subjecting refutation to any such unreasonably idealistic requirement, the folklore version of the Popper principle effectively reduces both confirmation and refutation to meaninglessness. In such a tendentious system … the Popper principle is not so much fallacious as, in the end, effectively empty.

While Carter objected to the norms of falsificationism, or what scientists generally but often wrongly took to be the norms, to critics of the anthropic principle the conflict indicated that Carter's principle was of a philosophical and not a scientific nature. Thus the prominent cosmologist Alexander Vilenkin, an advocate of eternal inflation and the multiverse, dismisses the anthropic principle as unscientific. In a popular book, he says: "The philosopher Karl Popper has argued that any statement that cannot be falsified cannot be scientific. This criterion, which has been generally adopted by physicists, seems to imply that anthropic explanations of the fine-tuning are not scientific" (Vilenkin 2006, 134).

3.3 *The Multiverse: Physics or Metaphysics?*

The idea of multiple universes or a "multiverse" goes far back in time, but in its modern version it dates from the 1990s. The general idea that there are numerous other universes in addition to the one we observe, and that most of the other universes differ entirely from ours with respect to the laws of physics and much more, was inspired principally by the many-worlds interpretation of quantum mechanics, string theory, the anthropic principle, and eternal inflation theories. For the last two decades the multiverse has been hotly debated, they main question being whether it is a scientific or a metaphysical idea (Carr 2007). In this ongoing discussion Popper and his demarcation criterion appear frequently and understandably so. After all, most versions of the multiverse operate with universes causally disconnected from ours and hence unobservable even in principle. How can such a claim be tested? Can it ever be falsified?

When confronted with such questions related to Popper's criteria, some multiverse physicists maintain that "the multiverse remains within the realm of Popperian science" (Barrau 2007). They agree that testability is a necessary epistemic standard also for cosmological theories but typically have in mind non-empirical testability such as mathematical consistency checks; or they point out that multiverse theories yield testable predictions in the form of probability distributions, say that the cosmological constant lies in a certain range. On the top of that it is sometimes argued that the multiverse is ultimately a consequence of the exceedingly well-tested theory of quantum mechanics.

Other physicists in favour of the multiverse, such as Leonard Susskind and Sean Carroll, simply dismiss Popper's standards of science as inadequate and irrelevant armchair philosophy. They claim that the criterion of falsifiability is far too blunt an instrument to discriminate between sound and unsound science in a field as advanced as theoretical cosmology. Susskind (2006, 196) has no respect at all for Popper and the "Popperazzi" who pontificate his principles of science. "As for rigid philosophical rules," Susskind states, "it would be the height of stupidity to dismiss a possibility just because it breaks some philosopher's dictum about falsifiability." More generally he and some other physicists insist that the issue of the defining criteria of science belongs to working scientists and not to philosophers.

Critics of the multiverse often invoke the authority of Popper to warn of the dangers of non-falsifiable theories. For example, George Ellis has at several occasions pointed out that multiverse physics fails catastrophically when confronted with Popperian standards. According to Popper (1963, 36), a good scientific theory forbids certain things to happen, but multiverse theories are extravagantly non-prohibitive. Whatever the prediction there is no doubt it will be confirmed at one of the zillions of universes making up the multiverse. Another prominent critic of string theory and the multiverse, the Canadian theorist Lee Smolin, proudly declares himself a Popperazo. In this respect there is more than a little similarity between him and Bondi. Almost repeating what Bondi said half a century earlier, Smolin (2007, 323) explains:

> According to Popper, a theory is falsifiable if one can derive from it unambiguous predictions for practical experiments, such that – were contrary results seen – at least one premise of the theory would have been proven not true. … Scientists have an ethical imperative to consider only falsifiable theories as possible explanations of natural phenomena.

Popper's analysis of the nature of science is not only part of the multiverse controversy but also of some other controversial issues related to the ΛCDM consensus theory. To the American astrophysicist Michael Turner (2001), an early advocate of inflation theory and ΛCDM, the new theory was appealing from a methodological point of view: "With its unidentified dark matter and mysterious dark energy, it is currently very much out on a limb. According to Karl Popper that's what strong theories do! … Inflation + cold dark matter is bold and testable."

Two decades later the situation looked different, at least to some astronomers and astrophysicists. The problem of cold dark matter, or the apparent absence of dark matter particles, has long been recognized. Currently a minority of astronomers challenges the belief in dark matter, arguing that it should be replaced by the MOND theory (Modified Newtonian Dynamics) going back to the early 1980s. Referring to Popper, Robert Sanders (2013, 168) concludes that whereas MOND is "inherently falsifiable," the cold dark matter concept is "fundamentally not falsifiable." More recently David Merritt (2017) has investigated the situation in relation to Popper's view of conventionalism. The issue is dealt with in detail by Merritt and Anastasiia Lazutkina in their contributions to the present volume.

3.4 According to Popper

The remarkable visibility of Popper in modern cosmology should be understood in its proper context, which is rhetorical and sociological rather than scientific. References to his views about science, sometimes explicit and at other times implicit, are usually boiled down to a much simplified version of the criterion of falsifiability. Such references are rare in research papers but appear frequently in popular books and articles, website discussions and reviews aimed at a general audience. Moreover, they mostly appear in connection with controversies between rival models and concepts, from the steady-state theory in the 1950s to the dark matter problem and the multiverse more than fifty years later.

As pointed out by Sovacool (2005), Kragh (2013) and Pigliucci (2019), when cosmologists and other scientists comment on the faults and merits of Popper's philosophy of science, it is often in a primitive folklore version that bears little resemblance to what Popper actually thought and wrote. The phrase "according to Popper" is typically followed by the claim that a good scientific theory must be falsifiable or even the absurd claim that if a theory is falsifiable it is also scientific. There is little doubt that few of the scientists, whether they are for or against Popperian criteria of science, have taken the care to study Popper. Perhaps they have looked into some of his works, but in other cases they rely on casual discussions with colleagues or what they can easily find on the internet.

Popper—the authentic Popper, that is—was not the naïve falsificationist as he is sometimes portrayed. He did indeed argue that falsifiability is a necessary condition for a theory being scientific, but he never suggested that it was a sufficient condition. Popper rated falsification higher than verification, and yet he did not deny that the latter plays an important role in the construction and evaluation of theories. Nor did he defend the notion of instant and definitive disproof. On the contrary, in *The Logic of Scientific Discovery* and elsewhere Popper (1959, 50) pointed out that in the practice of science "no conclusive disproof of a theory can ever be produced." As he stated in his autobiography, it will sometimes be rational to keep a theory alive even if it is refuted by convincing experiments. Ad hoc hypotheses and other means of protection reduce the degree of falsifiability but nonetheless have a legitimate place in science (see also the comprehensive and critical analysis in Grünbaum 1976).

"Logically speaking falsifiability, as testability, cannot be regarded as a very sharp criterion," Popper admitted (Schilpp 1974, 32). He warned against "supersensitivity with respect to refuting criticism" (Schilpp 1974, 984) and stressed that science is an intellectual activity based on disciplined conjectures:

> He who gives up his theory too easily in the face of apparent refutations will never discover the possibilities inherent in his theory. *There is room in science for debate*: for attack and therefore also for defence. … As always, science is conjecture. You have to conjecture when to stop defending a favourite theory, and when to try a new one.

These remarks should suffice to demonstrate that when cosmologists and other scientists refer to Popper and his ideas, their comments have in many cases no basis in what he actually said.

4 Conclusion

Although Popper had little to say about the physical cosmology which emerged in the 1950s and eventually resulted in the big-bang standard model, he followed the development. It seems that he was somewhat sceptical with regard to the grand project of a science of the universe, and at the end of his life he expressed his disillusion with the "intolerable" methods of modern cosmology. He dismissed the big-bang theory as essentially unscientific because it, as he saw it, violated the falsifiability criterion on which his philosophy of science rested.

The well-documented influence that Popper's views have had and still have on some branches of cosmology derives from his general criteria of good science and not from his scattered remarks concerning the scientific status of cosmology. In controversial areas such as the multiverse and dark matter Popperian standards are often discussed, if mostly in popular and general contexts. There is little doubt that Popper, or the shadow of Popper, is much more visible in modern physical cosmology than other philosophers of science. It has been argued that the falsifiability criterion has itself been falsified and that it fares poorly if compared to how scientists work and think (Hansson 2006). Perhaps so, but in some areas of physics and cosmology the criterion is very much alive.

References

Barrau, A.: Physics in the multiverse. Cern Courier **47**(10), 13–17 (2007)

Bondi, H.: Science and the structure of the universe. Mem. Proc. Manchester Lit. Phil. Soc. **101**, 58–71 (1958)

Bondi, H., et al. (eds.): Rival theories of cosmology. Oxford University Press, Oxford (1960)

Bondi, H.: Some philosophical problems in cosmology. In: Mace, C.A. (ed.) British Philosophy in the Mid-Century, pp. 393–400. Allen & Unwin, London (1966)

Carr, B.: Universe or multiverse?. Cambridge University Press, Cambridge (2007)

Carter, B.: The anthropic selection principle and the ultra-Darwinian synthesis. In: Bertola, F., Curi, U. (eds.) The Anthropic Principle, pp. 33–66. Cambridge University Press, Cambridge (1993)

Davies, J.T.: The age of the universe. Brit. J. Phil. Sci. **5**, 191–202 (1954)

Grünbaum, A.: Ad hoc auxiliary hypotheses and falsificationism. Brit. J. Phil. Sci. **27**, 329–362 (1976)

Hansson, S.O.: Falsificationism falsified. Found. Sci. **11**, 275–286 (2006)

Hawking, S.: A brief history of time: from the big bang to black holes. Bantam Books, New York (1989)
Hoyle, F., Burbidge, G., Narlikar, J.: A different approach to cosmology. Cambridge University Press, Cambridge (2000)
Keating, B.G.: Shut up and measure. https://www.edge.org/conversation/brian_g_keating-shut-up-and-measure (2017)
Kragh, H.: Cosmology and controversy: the historical development of two theories of the universe. Princeton University Press, Princeton (1996)
Kragh, H.: Karl Popper on physical cosmologies. J. Hist. Astron. **43**, 347–350 (2012)
Kragh, H.: "The most philosophically important of all the sciences": Karl Popper and physical cosmology. Persp. Sci. **21**, 325–357 (2013)
Kragh, H.: Alternative cosmological theories. In: Kragh, H., Longair, M. (eds.) The Oxford Handbook of the History of Modern Cosmology, pp. 120–161. Oxford University Press, Oxford (2019)
Merritt, D.: Cosmology and convention. Stud. Hist. Phil. Mod. Phys. **57**, 41–52 (2017)
Narlikar, J.: An introduction to cosmology. Cambridge University Press, Cambridge (2002)
Pigliucci, M.: Philosophy of science and the string wars: a view from the outside. In: Dardashti, R., Dawid, R., Thébault, K. (eds.) Why Trust a Theory? Epistemology of Fundamental Physics, pp. 84–98. Cambridge University Press, Cambridge (2019)
Popper, K.R.: Interpretations of nebular red-shifts. Nature **145**, 69–70 (1940)
Popper, K.R.: The logic of scientific discovery. Basic Books, New York (1959)
Popper, K.R.: Conjectures and refutations. Routledge, New York (1963)
Popper, K.R.: Time's arrow and feeding on negentropy. Nature **213**, 233–234 (1967)
Popper, K.R.: On the possibility of an infinite past: a reply to whitrow. Brit. J. Phil. Sci. **29**, 47–48 (1978)
Popper, K.R.: The open universe: an argument for indeterminism. Routledge, London (1982)
Popper, K.R.: In search of a better world: lectures and essays from thirty years. Routledge, London (1994)
Sanders, R.: The dark matter problem: a historical perspective. Cambridge University Press, Cambridge (2013)
Schilpp, P.A. (ed.): The philosophy of Karl Popper. Open Court Publishing House, La Salle, IL (1974)
Schrödinger, E.: Nature of the nebular red-shift. Nature **144**, 593 (1939)
Smolin, L.: Scientific alternatives to the anthropic principle. In: Carr, B. (ed.) Universe or Multiverse?, pp. 326–366. Cambridge University Press, Cambridge (2007)
Sovacool, B.: Falsification and demarcation in astronomy and cosmology. Bull. Sci. Tech. Soc. **25**, 53–62 (2005)
Steinhardt, P.: The endless universe: a brief introduction. Proc. Am. Phil. Soc. **148**, 464–470 (2004)
Susskind, L.: The cosmic landscape: string theory and the illusion of intelligent design. Little Brown and Company, New York (2006)
Toulmin, S.: The return to cosmology. University of California Press, Berkeley (1982)
Turner, M.: A sober assessment of cosmology at the new millennium. Publ. Astron. Soc. Pacific **113**, 653–657 (2001)
Vilenkin, A.: Many worlds in one: the search for other universes. Hill and Wang, New York (2006)
Weston Smith, M.: Beating the odds: the life and times of E.A. Milne. Imperial College Press, Singapore (2013)
Whitrow, G.J. (ed.): Einstein: the man and his achievements. BBC, London (1967)
Whitrow, G.J., Bondi, H.: Is physical cosmology a science? Brit. J. Phil. Sci. **4**, 271–283 (1954)

Helge Kragh is Professor Emeritus at the Niels Bohr Institute, University of Copenhagen and had previously been a professor of history of science and technology at Aarhus University, Denmark, and elsewhere. His principal research is in the history of post-1850 physical science,

including chemistry, astronomy and cosmology. He has also worked on the science-religion relationship. Among his books are Cosmology and Controversy (Princeton, 1996), Quantum Generations (Princeton, 1999), Higher Speculations (Oxford, 2011) and Niels Bohr and the Quantum Atom (Oxford, 2012). E-mail: helge.kragh@nbi.ku.dk

MOND and Methodology

David Merritt

1 .

Karl Popper, in his *Realism and the Aim of Science* (1983, p. 234), identified two main attitudes with respect to the testing of scientific theories:

> (a) The uncritical or verificationist attitude: one looks out for 'verification' or 'confirmation' or 'instantiation', and one finds it, as a rule. Every observed 'instance' of the theory is thought to 'confirm' the theory.
>
> (b) The critical attitude, or falsificationist attitude: one looks for falsification, or for counter-instances. Only if the most conscientious search for counter-instances does not succeed may we speak of a corroboration of the theory.

Attitude (b) is, of course, the attitude that Popper endorsed. A *critical* scientist denies that scientific theories are verifiable. She asserts that theories are to be judged on the basis of how well they stand up to critical appraisal—to sincere attempts at refutation.

In Popper's view, accommodating a theory to known experimental or observational results does not corroborate the theory, since "it is always possible to produce a theory to fit any given set of explicanda" (Popper 1963, pp. 241–2); or as Zahar (1973, p. 103) expressed it, "theories can always be cleverly engineered to yield the known facts." Corroboration occurs only when the theory survives an attempted falsification: that is: when it predicts a previously unknown fact and that prediction is subsequently confirmed. In the words of Imre Lakatos (1970, p. 38), "the only relevant evidence is the evidence anticipated by a theory."

There currently exist at least two, viable, cosmological theories: the standard, or 'concordance,' or ΛCDM, model; and an alternative theory, the foundational postulates of which were published by Mordehai Milgrom in 1983. The standard model assumes the correctness of Einstein's theory of gravity and motion (or of Newton's, in the appropriate regimes) and deals with anomalies via postulates relating to 'dark matter' and 'dark energy,' among others. Milgrom's theory includes no dark matter (or at least, does not require it). Observations that are explained in the standard

---

D. Merritt (✉)
Rochester, NY, USA
e-mail: david.r.merritt@gmail.com

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_5

model by invoking dark matter are explained in Milgrom's theory by postulating a modification to Newton's (or Einstein's) laws of gravity and motion.

Both the standard model, and Milgrom's theory—the latter is often called 'MOND', for MOdified Newtonian Dynamics'—can point to successes and failures (McGaugh 2015). But *only Milgrom's theory has repeatedly made novel predictions that were subsequently found to be correct* (Merritt 2020). Successes of the standard model have been, almost without exception, successes of *accommodation*: the theory has been adjusted, or augmented, or re-interpreted in order to bring its predictions[1] in alignment with new observational facts, most of which constituted problems for the theory when they first came to light. In many cases, those new facts *were* predicted in advance by Milgromian researchers; they were 'unexpected' only from the standpoint of standard-model researchers.

Dark matter has never been detected in any experiment that a particle physicist would consider decisive (Liu et al. 2017; Ko 2018; Kisslinger and Das 2019) and in the absence of such a detection the existence of dark matter remains an unconfirmed hypothesis. But there is a related question that *is* capable of being decisively answered: to which methodological 'school' do researchers in the two camps typically belong? The published record betrays a distinct, and profound, difference: Milgromian researchers have adhered closely to Popper's methodology, standard-model cosmologists have not. I argue that these methodological differences render pointless any discussion of the comparative 'truthlikeness' or 'verisimilitude' of the theories since the two groups of scientists achieve correspondence with data in fundamentally different ways.

2 .

The typical response of a scientist to a falsifying instance—an experiment or observation that contradicts a theory— is not to discard the theory. Scientists are more likely to retain the theory and ignore the counterexample (Kuhn 1962; Lakatos 1973). If the refutation is persistent or compelling, the scientist may decide to tack an additional hypothesis onto the theory, one that targets the anomaly and 'explains' it. Ptolemy's 'equants' came about in this way, as did the postulates in the standard cosmological model about 'dark matter' and 'dark energy' (Merritt 2017).

Karl Popper's name is most often associated with his famous demarcation criterion: the idea that scientific hypotheses have the quality of being falsifiable, that is, vulnerable to experimental testing. But Popper was quite aware that scientists do not always walk away from a theory just because it has failed a test. In fact, he wrote at length, and with considerable insight, about the methodology that scientists should follow when modifying a theory in response to refutations.

---

[1] Throughout this chapter I use the term 'prediction' in the same way that Popper does, to describe a statement that follows logically (deductively) from a theory (see e.g. item 28 in Table 1); it comprises 'retrodiction' and 'explanation' (e.g. Popper (1957, p. 133)).

Popper's methodological guidelines were intended to preserve falsifiability: "*Only with reference to the methods applied* to a theoretical system is it at all possible to decide whether we are dealing with a conventionalist or an empirical theory" (1959, p. 82). He understood that, logically, falsification could always be evaded by conventionalist maneuvers: by ad hoc changes that simply target the anomaly. Thus Popper required that a modified theory should do more than simply explain the experimental results that brought down the previous theory. The new theory should have more content, and it should only be accepted if it passes some new tests, among other requirements.

Many of Popper's methodological rules appeared in 1934, in *Logik der Forschung*. Others can be found scattered through later writings, including *Conjectures and Refutations* (1963), *Realism and the Aim of Science* (1983), and *The Open Society and its Enemies* (1945) among others. Jarvie (2001) lists fifteen rules; Keuth (2005) finds twelve, only six of which appear in Jarvie's list; and Johansson (1975) compiles over twenty, including rules for the social sciences, the latter mostly from *The Poverty of Historicism* (1957).

Table 1 presents a concatenation of these three lists. I have omitted the rules pertaining to social science and to probability statements, and I include two additional rules: no. 3 (from Popper 1959, p. 253) and no. 8 (from Popper 1963, p. 38).

The first rule in Table 1 was called by Popper (1959, p. 33) the "supreme rule," that is, "a rule of a higher type. It is the rule which says that the other rules of scientific procedure must be designed in such a way that they do not protect any statement in science against falsification". (Keuth calls this the "meta-rule".)

The second rule ("The game of science is, in principle, without end") expresses Popper's commitment to fallibilism: the acknowledgement that we can never be certain of the correctness of our theories, therefore we can never stop testing them.

Rules no. 3–6 enjoin the scientist to search for causal and universal laws as explanations for observed events. As David Miller (1994, pp. 26–7) has emphasized, Popper is not implying here any "metaphysical assumption concerning the immutability or order of nature," since such an assumption is amenable to testing and may be found to be false. Rather he is proposing the methodological rule: search for spatio-temporally invariant laws, even though the search may turn out to be unsuccessful.

Rules no. 8–13 forbid conventionalist stratagems, that is, adjustments intended to protect a theory from falsification. In *The Logic of Scientific Discovery*, Popper highlighted four such "immunizing" techniques and rules nos. 10–13 target each in turn. Jarvie (2001) separates rule no. 13 into two: admonishing scientists, when faced with a refutation, not to arbitrarily reject either an experimental result or the theoretical derivation that conflicts with it.[2]

In addition to forbidding conventionalism, Popper proposed a number of other rules that are relevant to theory change—that is: to a situation in which a theory is

---

[2] Both Johansson (1975) and Jarvie (2001) note that the wording of the first part of rule no. 13 is confusing. Johansson (p. 58) suggests that Popper meant to write "Inter-subjectively testable *theories*"; Jarvie (p. 59) suggests that "What is plainly intended is a presumption that inter-subjectively testable experimental work be accepted." I find Jarvie's suggestion to be the more convincing.

Table 1 Popper's methodological rules

#	Rule	Source
1	The other rules of scientific procedure must be designed in such a way that they do not protect any statement in science against falsification	J [SR], K [MR]
2	The game of science is, in principle, without end. He who decides one day that scientific statements do not call for any further test, and that they can be regarded as finally verified, retires from the game	J1, K1
3	It is part of our *definition* of natural laws if we postulate that they are to be invariant with respect to space and time; and also if we postulate that they are to have no exceptions	M1
4	We are not to abandon the search for universal laws and for a coherent theoretical system, nor ever give up our attempts to explain causally any kind of event we can describe	J3, K4
5	*Never …explain physical effects, i.e. reproducible regularities, as accumulations of accidents*	J12
6	Regard natural laws as synthetic and strictly universal statements	3i
7	Only such statements may be introduced in science as are inter-subjectively testable	1i, K3
8	*Criteria of refutation* have to be laid down beforehand: it must be agreed which observable situations, if actually observed, mean that the theory is refuted	M2
9	In the case of a threat to our system, we will not save it by any kind of *conventionalist stratagem*	K6
10	Adopt a rule not to use undefined concepts as if they were implicitly defined	2i, J4
11	Only those [auxiliary hypotheses] are acceptable whose introduction does not diminish the degree of falsifiability or testability of the system in question, but, on the contrary, increases it	2ii, J5, K8
12	We shall forbid *surreptitious* alterations of usage	2iii, J6, K9
13	Inter-subjectively testable experiments are either to be accepted, or to be rejected in the light of counter-experiments. The bare appeal to logical derivations to be discovered in the future can be disregarded	2iv, J7-8
14	Auxiliary hypotheses shall be used as sparingly as possible	2v
15	The number of our axioms – of our most fundamental hypotheses – should be kept down	3iii
16	The new theory should proceed from some *simple, new, and powerful, unifying idea* about some connection or relation (such as gravitational attraction) between hitherto unconnected things (such as planets and apples) or facts (such as inertial and gravitational mass) or new 'theoretical entities' (such as field and particles)	3v
17	Any new system of hypotheses should yield, or explain, the old, corroborated, regularities	3iv

(continued)

Table 1 (continued)

#	Rule	Source
18	Those theories should be given preference which can be most severely tested …equivalent to a rule favouring theories with the highest possible empirical content	3ii, J11
19	We require that the new theory should be *independently testable*	3vi
20	We shall take it [the theory] as falsified only if we discover a *reproducible effect* which refutes the theory. In other words, we only accept the falsification if a low-level empirical hypothesis which describes such an effect is proposed and corroborated	4i
21	We should not accept *stray basic statements* – i.e. logically disconnected ones – but …we should accept basic statements in the course of testing *theories*; of raising searching questions about these theories, to be answered by the acceptance of basic statements	4ii, J10
22	A theory is to be accorded a positive degree of corroboration if it is compatible with the accepted basic statements and if, in addition, a non-empty sub-class of these basic statements is [– – –] accepted as the results of sincere attempts to refute the theory	5i
23	It is not so much the number of corroborating instances which determines the degree of corroboration as *the severity of the various tests* to which the hypothesis in question can be, and has been, subjected	5ii
24	we shall not continue to accord a positive degree of corroboration to a theory which has been falsified by an inter-subjectively testable experiment	5iii, K12
25	We require that the [new] theory should pass some new, and severe, tests	5iv
26 27	We choose the theory which best holds its own in competition with other theories; the one which, by natural selection, proves itself the fittest to survive. This will be the one which not only *has hitherto stood up to the severest tests*, but the one which *is also testable in the most rigorous way* Whenever we find that a system has been rescued by a conventionalist stratagem, we shall test it afresh, and reject it, as circumstances may require	K10 K7
28	With the help of other statements, previously accepted, certain singular statements – which we may call ‘predictions’ – are deduced from the [new] theory; especially predictions that are easily testable or applicable. From among these statements, those are selected which are not derivable from the current theory, and more especially those which the current theory contradicts	K5
29	a theory which has been well corroborated can only be superseded by one of a higher level of universality; that is by a theory which is better testable and which, in addition, *contains* the old, well corroborated theory – or at least a good approximation to it	K11

(continued)

Table 1 (continued)

#	Rule	Source
30	Once a hypothesis has been proposed and tested, and has proved its mettle, it may not be allowed to drop out without 'good reason'. A 'good reason' may be, for instance: replacement of the hypothesis by another which is better testable; or the falsification of one of the consequences of the hypothesis	J2, K2
31	after having produced some criticism of a rival theory, we should always make a serious attempt to apply this or a similar criticism to our own theory	J9

'2i' indicates the i'th rule from group 2 of Johansson (1975) and similarly for J (= Jarvie 2001) and K (= Keuth 2005); reference to the works by Popper in which the rules first appeared can be found by consulting those authors. Rules no. 3 and 8, marked 'M', do not appear in any of the three lists and references are given in the text

modified in response to a refutation. Rules nos. 14–19, 25 and 26 together imply the following: In explaining the observations that brought down a falsified theory, a new theory should conserve the explanatory successes of the old theory (rule no. 17); it should do so in a way that maximizes falsifiability/new content/boldness (nos. 14–16, 18); and at least some of the modified theory's novel content should be experimentally corroborated (nos. 25, 26). Of course, as stated, rule no. 25—that "the theory should pass some new, and severe, tests"—is not quite a *methodological* rule, since, as Lakatos (1968, p. 388) noted, "It is up to us to devise bold theories; it is up to Nature whether to corroborate or to refute them." But it is reasonable to recast the rule as a methodological one, e.g., we *accept* a new theory only if it passes some new, and severe, tests.

Most scientists would probably agree with Popper about the privileged status of confirmed, novel predictions. For instance, Gottfried Leibniz wrote that "Those hypotheses deserve the highest praise …by whose aid predictions can be made, even about phenomena or observations which have not been tested before" (Leibniz 1678). Similar statements can be found in writings of John Herschel, William Whewell, Henri Poincaré, Charles Peirce, Norbert Campbell and others. But Popper makes a stronger claim. Not only is it *impressive* when a theory correctly predicts a previously unknown fact. Popper is arguing that the confirmation of a novel prediction is the *only sort of evidence that counts*.

What was the basis for this claim? The starting point is the fallacy of induction: the logical impossibility of generalizing from discrete instances to a general rule. Even an incorrect theory can make correct predictions, and one can always accommodate a finite set of data to an infinite number of theories.

It is tempting to believe that a successful prediction lends support to a theory, but this belief flies in the face of the 'paradoxes of confirmation' (Hosiasson-Lindenbaum1940; Hempel 1945). It is easy to show that, from a purely logical standpoint, a universal hypothesis is supported by anything that does not contradict it; the only sort of observation that fails to support a hypothesis is one that disproves it. Thus: my observation of a red fire truck outside my window confirms the standard model of cosmology precisely as much (or as little) as an observation of the cosmic microwave background—so long as that model does not forbid the existence of red fire trucks. As Popper (1983, p. 235) put it: "Thus an observed white swan will, for the verificationist, support the theory that all swans are white; and if he is consistent (like Hempel), then he will say that an observed black cormorant also supports the theory that all swans are white." The 'paradoxes' of confirmation are a straightforward consequence of the fallacy of induction.

Correspondence of data with theory, of itself, counts for little; one needs to find a sharper criterion to separate the evidentially relevant wheat from the chaff.

Popper's 'positive theory of corroboration' (1959, pp. 265–73; 1983, pp. 230–61) derives from three premises.[3] From the paradox of confirmation it follows (as just discussed) that it would be a mistake to consider an observation as supporting a hypothesis simply because it is consistent with that hypothesis. The second premise was Popper's belief that, while degree of corroboration "may at first sight look like a probability …[it] exhibits properties incompatible with the rules of the probability calculus" (1983, p. 232). For instance, the testability of a theory, and therefore its potential for corroboration, increases with its informative content, and therefore with its *im*probability. And third was Popper's insistence that a goal of science must always be toward theories with *greater* informative content—that is: toward theories of lower probability. Popper (1983, p. 222) argued that an inductivist (he singled out Carnap) will always try to maximize the probability of a theory given the evidence and so will always be led to theories that go as little as possible beyond the evidence. Whereas scientists, he said, "invariably prefer a highly testable theory whose content goes far beyond all observed evidence to an ad hoc hypothesis, designed to explain just this evidence, and little beyond it, even though the latter must always be more probable than the former" (Popper 1983, p. 256).

Popper (1972, p. 71) noted that "Knowledge never begins from nothing, but always from some background knowledge." He defined background knowledge, B, as the set of assumptions that are accepted (perhaps only tentatively) when a new hypothesis is tested, and argued that "what is interesting in a new conjecture a is, in the first instance, the relative content a, B; that is to say, that part of the content of a which goes beyond B" (Popper 1972, p. 49).

---

[3] Philosophers who reject some or all of these premises will sometimes nevertheless embrace the *conclusions* that Popper derived from them. For instance, Psillos (1999), who makes no secret of his inductivist leanings, or of his admiration for Carnap, writes (pp. 105 and 173) "we should not accept a hypothesis merely on the basis that it entails the evidence, if that hypothesis is the product of an ad hoc manoeuvre …The notion of empirical success that realists are happy with is such that it includes the generation of novel predictions which are in principle testable." Those sentences could just as easily have been written by Popper (cf. rules no. 7, 9 and 19 from Table 1). Niiniluoto (2018, p. 117) similarly suggests as an "acceptance rule" for an inductive inference that it "should be independently testable, i.e. it should either explain some old evidence or be successful in serious new tests …the best hypothesis is one with both explanatory and predictive power.".

Based on these arguments, Popper was led to reformulate the question ‘Does an observation E support a hypothesis H?’ as ‘Does E support H in the presence of background knowledge B?’

Simply requiring that E follow from the conjunction of H and B is insufficient, since this condition may be satisfied if E follows from B alone. Nor is it enough to demand that E does *not* follow from B alone. Popper gave as an example the failure of James Challis to discovered Neptune, even though he was the first to observe the planet near to its predicted location: “The presence of *some* unknown star of eighth magnitude, close to the calculated place, was in itself quite probable on his background knowledge and therefore did not appear significant to him” (Popper 1983, p. 237).

These considerations led Popper (1983, p. 239) to propose that evidence E supports hypothesis H given background knowledge B if both:

1. E follows from the conjunction of H and B
2. E is improbable on the background knowledge alone.

Popper noted that saying that E is improbable based on B alone is similar to saying that H is a bold hypothesis – that it makes claims that go far beyond the background knowledge; and therefore that it has high empirical content. Elsewhere, Popper had defined a “severe test” in essentially the same way. Thus Popper’s condition for corroboration can be stated as: A theory is corroborated when it survives a severe test: a concerted attempt at falsification.[4] The more novel a test—the more unlikely the prediction in the light of existing knowledge—the riskier it is, and the greater the degree of corroboration if the prediction is confirmed.

3 .

Subsequent authors—while not objecting to Popper’s basic reasoning—have argued for different, or broader, definitions of what constitutes a ‘novel prediction’ or evidential support. Zahar (1973) noted that Popper’s criterion, which recognizes only observations made after a theory was formulated, excludes some well-known examples from history. For instance, the Balmer series of hydrogen was known before Bohr published his postulates in 1913; Kepler’s laws were known to Newton; Einstein knew of the anomalous precession of Mercury’s orbit. What matters, Zahar argued, is not the chronology so much as whether a fact “belong[s] to the problem-situation which governed the construction of the hypothesis”—i.e., whether the theory was designed to explain the fact. On this view, a ‘novel fact’ is one that a theory was not specifically designed to explain.

Of course, one does not always know what background knowledge was in the mind of the theorist who designed the theory. But there is one—rather common—

---

[4] Miller (1994, p. 106): “Sitting around complacently with a well-meant resolve to accept any refutations that happen to arise is a caricature of genuine falsificationism.”.

circumstance in which it is obvious that background knowledge is being used in this way. That is when the theory contains unspecified parameters, and the parameters are determined from experimental or observational data. In Zahar's (1973, pp. 102–3) words:

> Consider the following situation. We are given a set of facts and a theory T [$\lambda_1, \ldots, \lambda_m$] which contains an appropriate number of parameters. Very often the parameters can be adjusted so as to yield a theory T* which 'explains' the given facts …In such a case we should certainly say that the facts provide little or no evidential support for the theory, since *the theory was specifically designed to deal with the facts*.

Zahar is not claiming here that there is anything illegitimate about determining a theory's parameters from data. Rather, he is arguing that data that are used to set the parameters of a theory do not *corroborate* the theory; they only *complete* the theory; and in so doing they have lost their evidentiary value. John Worrall summarized this condition more succinctly: "one can't use the same fact twice: once in the construction of a theory and then again in its support" (Worrall 1978, p. 48).[5]

In practice, the situation may not be quite as clear-cut as Zahar's argument suggests. There may be different data sets, or combinations of data sets, that a theorist can use when determining a theory's parameters and it may not be obvious which data should be assigned to the 'background knowledge' and which data can be considered evidentially relevant. It is also possible that the data are not reproducible for *any* choice of a theory's parameters, and if so they would expose the theory to potential falsification.[6]

But consider the following special case. Suppose that the theory contains just one unspecified parameter and that its value is formally over-determined by the data: that is, that there are a number of independent data sets from which the parameter can be determined with comparable precision. In that case, whichever data set is chosen to determine the parameter (it does not much matter which) becomes part of the background knowledge; the newly-determined parameter can then be inserted into the theory and the now-completed theory can be used to make predictions which can be tested against the other data sets. Those tests are novel according to Zahar's

---

[5] Worrall (1985, p. 313) argues further "that when one theory has accounted for a set of facts by parameter-adjustment, while a rival accounts for the same facts directly and without contrivance, then the rival does, but the first does not, derive support from those facts." Interpreted broadly, Worrall's argument would imply that no standard-model explanation of any fact correctly predicted by Milgrom's theory counts in favor of the standard model, since standard-model explanations of such facts always invoke a multitude of adjustable parameters or auxiliary hypotheses not required by Milgrom's theory; some examples are discussed below.

[6] An example occurred in studies of the cosmic microwave background (CMB), but the response of standard-model cosmologists was simply to add more parameters. Early studies of the CMB (e.g. Jaffe et al. 2001) assumed a value $n = 1$ for the power-law index of the spectrum of initial density perturbations, but as the amount and quality of data increased, this value n began to be treated as a free parameter (e.g. Netterfield 2002) and later as a 'running index' (e.g. Spergel et al. 2007). In this way the model was "immunized" (Popper's expression) from falsification. I am aware of only one attempt to confront the CMB data with a testable prediction; the theory was Milgrom's and the prediction (McGaugh 1999) was confirmed (de Bernardis et al. 2002).

Table 2 Determinations of Milgrom's constant

Prediction	References	N_{galaxy}	a_0 (10^{-10} m s^{-2})
Baryonic Tully-Fisher relation ($a_0 G$)	Begeman et al. (1991)	10	1.21 ± 0.24
	Stark et al. (2009)	28	1.18
	+ Trachternach et al. (2009)	34	1.30
	McGaugh (2011)	47	1.24 ± 0.14
	Lelli et al. (2016a)	118	1.29 ± 0.06
Central surface density Relation (a_0/G)	Donato et al. (2009)	$\sim 10^3$	1.3
	Lelli et al. (2016b)	135	1.27 ± 0.05
Radial acceleration relation ($a_0/G \to a_0 G$)	Wu and Kroupa (2015)	74	1.21 ± 0.03
	McGaugh et al. (2016)	153	$1.20 \pm 0.02 \pm 0.24$
	Lelli et al. (2017)		

criterion, and if they are successful, the successes constitute corroboration of the underlying theory.

This is a good description of how Milgrom's constant a_0—the only undetermined parameter in his theory—is determined. Table 2 (adapted from Merritt 2020) demonstrates that a number of independent data sets, targeting three predictions of Milgrom's theory, yield comparable, and comparably accurate, estimates of a_0, approximately 1.2×10^{-8} cm s^{-2}. (In fact, as discussed in the next section, one can determine a_0 independently using any one of hundreds of existing galaxy rotation curves, although with less precision.) Having determined a_0 using any one of the data sets in Table 2, a scientist can insert that value into Milgrom's theory and make quantitative predictions that are testable using any of the other data sets listed there. As discussed in the next section, those predictions turn out to be successful; and since Zahar's criterion is satisfied for them, those successes can be said to corroborate Milgrom's theory.

The philosopher Losee (2004, p. 156; 2005, p. 166) uses the term 'convergence' to describe cases like this: in which a new constant of nature is determined, consistently, from a number of different kinds of data. Losee writes "I know of no plausible countercase in which convergence of this kind is achieved in the case of a transition judged not to be progressive on other grounds":

> The convergence of various determinations of the value of Avogadro's number on 6.02×10^{23} molecules/ gram molecular weight warrants as progressive the transition from theories of the macroscopic domain to the atomic-molecular theory of its microstructure. And the convergence of various determinations of the value of Planck's constant on 6.6×10^{-27} erg-sec warrants the transition from classical electromagnetic theory to the theory of the quantization of energy (Losee 2004, pp. 156–7).

Losee (who was unaware, apparently, of Milgrom's theory) adds "Unfortunately, opportunities to apply the convergence condition are rare within the history of science."

Does the standard cosmological model provide any opportunities for testing convergence? Indeed it does: the mean baryon[7] density, ρ_b, is a parameter that can be measured in a number of independent ways. It is traditional to express this quantity in terms of the dimensionless 'concordance' parameter Ω_b as

$$\rho_\mathrm{b} = \frac{3}{8\pi G}\Omega_b H^2 \tag{1}$$

with H the Hubble (expansion) parameter. The concordance value of $\Omega_b h^2$ is said to be 0.022 where $h \equiv H_0/100$ km s^{-2}. Prior to observations of the CMB in the early 2000s, the value of ρ_b was determined from two, quite different sorts of data: (*i*) the measured abundance of ^{7}Li in the atmospheres of Population II stars in the halo of the Milky Way, together with the equations of big-bang nucleosynthesis; and (*ii*) direct census of the density of matter in the local universe. Both techniques yielded (and continue to yield) $0.011 \lesssim \Omega_b h^2 \lesssim 0.016$ –consistent with each other, and roughly one-half of the current concordance value. Since about 2002, standard-model cosmologists have *excluded* these two data sets when determining the values of the parameters that define their 'concordance' model; the resulting discrepancies in the value of ρ_b are called by them the 'lithium problem' (e.g. Fields 2011) and the 'missing baryons problem' (e.g. Shull et al. 2012). Thus, the evolution of the standard cosmological model beginning around 2002 violated rules no. 17 and 29 in Table 1: it failed to conserve "the old, corroborated, regularities" of the model, namely, the convergence of measured values of ρ_b.

Neither the 'lithium problem' nor the 'missing baryons problem' exists from the standpoint of a Milgromian researcher, who is likely to prefer the value of ρ_b that was established prior to 2000.

4 .

Milgrom's theory is a response to an anomaly that arose in the 1970s in observations of disk galaxies. The speed, V, at which stars or gas clouds orbit at distance R about the galaxy center is predictable using Newton's laws of gravity and motion given the observed distribution of mass ('baryons') in the galaxy. The Newtonian prediction is often found to be reasonably correct near the centers of galaxies, i.e. $V_\mathrm{obs}(R) \approx V_\mathrm{Newton}(R)$. But at sufficiently large R, rotation curves become 'asymptotically flat':

---

[7] Standard-model cosmologists often use 'baryonic matter' to mean 'normal [i.e. non-dark] matter'. Milgromian researchers sometimes follow suit, even though, from their perspective, there is no need to distinguish between two sorts of matter.

the orbital speed tends to a constant value (different in different galaxies), $V_{\text{obs}}(R) \rightarrow V_\infty \gg V_{\text{Newton}}$.

In his first paper from 1983, Milgrom proposed a modification to Newton's laws of gravity and motion that targets, and explains, the asymptotic flatness of galaxy rotation curves. Milgrom's auxiliary hypothesis was presented in the form of three postulates, which were re-stated, in slightly different form, in two subsequent papers from the same year. For the sake of definiteness I will take the liberty of (re-)stating Milgrom's postulates as follows:

1. Newton's second law relating acceleration to gravitational force is asymptotically correct when applied to motion for which the gravitational acceleration is sufficiently large, but breaks down when the acceleration is sufficiently small.
2. In the limit of small gravitational accelerations, the acceleration of a test particle, in a symmetric and stationary gravitating system, is given by $(a/a_0)\,\boldsymbol{a} \approx \boldsymbol{g}_{\text{N}}$, where $\boldsymbol{g}_{\text{N}}$ is the conventional gravitational acceleration and a_0 ('Milgrom's constant') is a constant with the dimensions of acceleration.
3. The transition from the Newtonian regime to the low acceleration regime is determined by Milgrom's constant. The transition occurs within a range of accelerations of order a_0 around a_0.

Sufficiently far from the center of a galaxy, the Newtonian gravitational acceleration has magnitude $\left|\boldsymbol{g}_{\text{N}}\right| \approx GM_{\text{gal}}/r^2$, with M_{gal} the total mass of the galaxy and r the distance measured from the galaxy center. Milgrom's second postulate, $(a/a_0)\boldsymbol{a} \approx \boldsymbol{g}_{\text{N}}$, then implies

$$a \approx \left(a_0 G M_{\text{gal}}\right)^{1/2} r^{-1}. \tag{2}$$

Equating this expression with the centripetal acceleration of a test mass moving in a circular orbit of radius $r = R$, or V^2/R, yields

$$\frac{V^2}{R} = \frac{\left(a_0 G M_{\text{gal}}\right)^{1/2}}{R} \quad \text{i.e.} \quad V = \left(a_0 G M_{\text{gal}}\right)^{1/4}, \tag{3}$$

so that V is independent of R.

As Milgrom pointed out in the same three papers from 1983, his postulates can be used to generate additional, testable predictions. As we will see, most or all of these predictions have been observationally confirmed. But before discussing those results we can pause and take stock of how well Milgrom's proposed changes to Newton's laws accord with Popper's methodological rules:

Rules nos. 7, 11 and 19, which require that theory modifications be testable and that they result in increased empirical content, are (as just noted) clearly satisfied. Rule no. 8 ("*criteria of refutation* have to be laid down beforehand") is also clearly satisfied: it is obvious from Milgrom's 1983 papers that he viewed his predictions as having the potential to falsify the underlying theory.

Rules nos. 14 and 15, which counsel parsimony when adding axioms (three in this case), are arguably satisfied, as is rule no. 16 ("should proceed from some simple, new, and powerful" idea): a proposal that Newton's laws are incorrect is nothing if not “new, and powerful.” And Milgrom’s auxiliary hypotheses clearly preserve “old, corroborated, regularities” (rule no. 17): both in the high-acceleration regime ($a \gg a_0$) since the modified theory makes the same prediction as Newton’s laws; and also in the low-acceleration regime, since the only “regularity” that was known to exist in this regime ca. 1980 was the asymptotic flatness of rotation curves. (In other words: no ‘Kuhn losses’ here.) None of the other rules 1–24 (to the extent that they are applicable) is violated.

This brings us to rule no. 25, which demands that the modified theory “should pass some new, and severe, tests”—in other words, that (at least some of) its novel predictions should be experimentally corroborated. And here, Milgrom’s modified dynamics has performed not just adequately, but—by any reasonable standard—spectacularly. Here is a partial list of corroborated novel predictions:

1. A universal relation between asymptotic speed and total mass of a disk galaxy
2. A universal relation between the acceleration $\boldsymbol{a}$ at any point in a disk galaxy and the Newtonian gravitational acceleration $\boldsymbol{g}_{\mathrm{N}}$ due to the galaxy’s mass
3. A universal relation between the central surface densities of normal and ‘dark’ matter in galaxies
4. A predicted dependence of the rms, vertical velocity of stars on distance above or below the plane of the Milky Way galaxy
5. A relation between the mass of a gravitating system and the root-mean-square velocity of its components

Prediction no. 1 is just Eq. (3). The *observed* relation between galaxy mass and asymptotic rotation speed (Fig. 1a) is nowadays called the ‘baryonic Tully-Fisher relation’ or BTFR.[8] The *predicted* relation contains (like essentially all predictions from Milgrom’s theory) the unspecified constant a_0, ‘Milgrom’s constant.’ As noted earlier (cf. Table 2 and the accompanying discussion), there are many ways to determine a_0 from data but most astrophysicists consider the BTFR to be the ‘cleanest,’ that is, least subject to systematic errors. The novelty of Milgrom’s prediction can perhaps best be attested by the fact that no standard-model cosmologist had proposed (or, it appears, searched for) any such relation prior to 1983—no doubt in large part because, under the standard model, the asymptotic velocity is attributable almost entirely to the dark matter, not the ‘baryons.’

Prediction no. 2 is probably the most remarkable, at least from the standpoint of standard-model expectations. Milgrom (1983a) showed that the local acceleration a—accessible, in any disk galaxy, through $V^2(R)/R$—must be related to the gravitational acceleration g_{N} computed from the observed mass distribution under Newtonian gravity via a relation $a = f(g_{\mathrm{N}}/a_0)g_{\mathrm{N}}$ with $f(x)$ some as yet unspecified function of $x \equiv g_{\mathrm{N}}/a_0$. (Of course, in the purely Newtonian case, $f \equiv 1$.) Regardless of the functional form of f, Milgrom’s prediction was therefore that the acceleration will be a *universal function of the Newtonian prediction, hence of the baryonic*

---

[8] That rather baroque name is due to standard-model cosmologists; see e.g. Merritt (2020, Chap. 4) for the relevant history. Milgrom refers to his predicted relation by the much more apt name ‘mass–asymptotic speed relation.’.

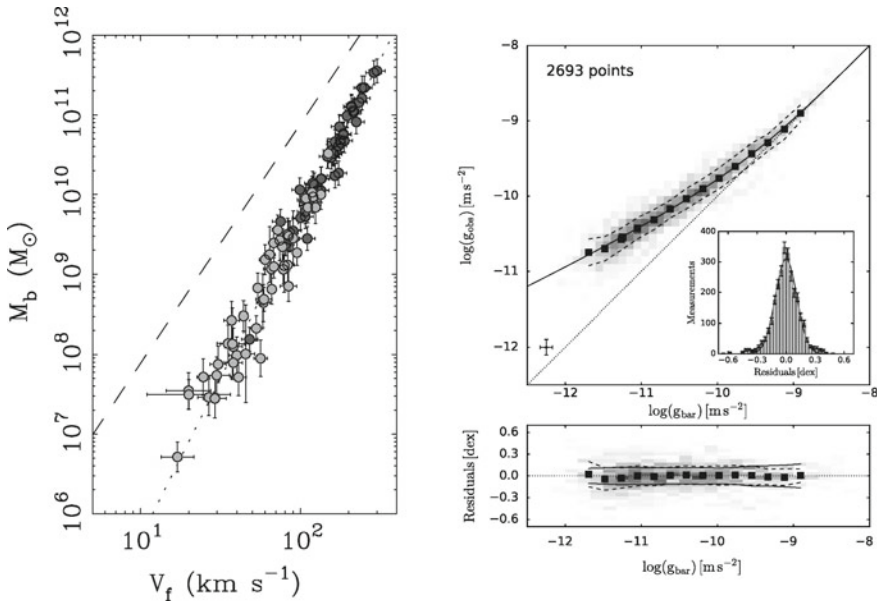


Fig. 1 Two confirmed, novel predictions of Milgrom's theory. *Left*: The baryonic Tully-Fisher relation (BTFR). Each point corresponds to a single disk galaxy. The vertical axis (M_b) is the summed mass in stars and gas (the subscript 'b' stands for 'baryonic,' i.e., 'non-dark'). The horizontal axis (V_f) is the outer, flat rotation velocity as inferred from 21 cm radio telescopic observations, what is called here 'V_∞'. Figure reprinted with permission from B. Famaey and S. S. McGaugh, "Modified Newtonian dynamics (MOND): observational phenomenology and relativistic extensions," *Living Reviews in Relativity*, 15, 2012, p. 20. *Right*: The radial acceleration relation (RAR), derived from rotation curve data of 153 galaxies. The vertical axis plots the observed acceleration, $g_{\rm obs} = V^2/r$, or what is called a in the text. The horizontal axis plots $g_{\rm bar} = |\partial\Phi/\partial r|$, or what is called $g_{\rm N}$ in the text. Figure reprinted with permission from McGaugh et al. (2016). Copyright (2016) by the American Physical Society

mass distribution, with a_0 as the only unspecified parameter. As noted by Milgrom (2016), this prediction can be expressed in a number of essentially equivalent ways: as a relation between a and $g_{\rm N}$, between a and $a/g_{\rm N}$, between $g_{\rm N}$ and $a/g_{\rm N}$ etc. And in fact, observational corroboration of the prediction exists in three forms: individual rotation curves; the 'mass discrepancy-acceleration relation' (MDAR), i. e. the relation between $a \equiv V^2/R$ and $a/g_{\rm N}$; and the 'radial-acceleration relation' (RAR), the relation between a and $g_{\rm N}$. Figure 1b compares the observed and predicted RAR.

Given data like those in Fig. 1b, one can simply 'read off' (modulo measurement uncertainties) the functional form of the relation between a and $g_{\rm N}$. The latter is sometimes called the 'transition function.' Various ad hoc, analytic forms have been proposed for the transition function; the curve plotted in Fig. 1b has a form first suggested by McGaugh (2008),

$$f(y) = \left[1 - \exp\left(-y^{1/2}\right)\right]^{-1}, \quad y \equiv \frac{a}{g_{\mathrm{N}}}. \tag{4}$$

Note the remarkable fact that Eq. (4) (to the extent that it is accurate) then allows one to predict the acceleration experienced by a test body orbiting *anywhere* in a galactic disk—not just in the 'asymptotic' regions of high or low acceleration! Thus, for instance, one can predict the rotation curve $V(R)$ for any single disk galaxy having a well-determined mass distribution and there are dozens of published examples of this kind (Milgrom 1988 being one of the earliest). Perhaps the most striking of these studies (e.g. de Blok and McGaugh 1997) are based on galaxies which a standard-model cosmologist would claim are 'dark matter dominated' everywhere—that is: for which the observed rotation speed greatly exceeds the Newtonian prediction at all positions. A standard-model cosmologist would predict that the rotation curve of such a galaxy is essentially independent of the (non-dark) matter distribution, and yet one finds, in every case, that the rotation curve is correctly predicted by the modified dynamics using only the observed density in stars and gas.

Given the background knowledge that existed ca. 1980, the proposal that the kinematics of any disk galaxy could be predicted, with high accuracy, from the observed distribution of normal matter alone was amazingly bold. There was simply no basis, under the standard model, for believing any such thing, and yet it turned out to be correct.

The remaining predictions nos. 3–5 and their observational corroboration are discussed in detail in Merritt (2020).

A number of Popper's remaining methodological rules 26–31 are expressed with reference to the competing theory, to which I now turn. In what follows I will restrict the discussion to predictions that follow from the non-relativistic versions of both theories; see Merritt (2020) for a discussion of the relativistic theories.

5 .

Standard-model cosmologists deal with the rotation curve anomaly by leaving Newton's laws intact and adding an auxiliary hypothesis[9]:

SCM1 In any galaxy or galactic system for which the observed motions are inconsistent with the predictions of Newton, the discrepancy is due to gravitational forces from dark matter in and around the galaxy(ies).

This auxiliary hypothesis has far less informative content—that is, it is far less *testable*—than Milgrom's. As we have seen, Milgrom's hypothesis allows one to

---

[9] Few cosmology textbooks acknowledge that the existence of dark matter is a postulate. Standard-model cosmologists take it for granted, apparently, that the existence of dark matter has been verified by rotation curve studies; e.g. Schneider (2015, p. 77): "The rotation curves of spiral galaxies are flat up to the maximum radius at which they can be measured; *spiral galaxies contain dark matter*" (italics his). Milgrom deserves credit for emphasizing that the existence of 'dark matter' is a postulate of the standard model and not a confirmed fact.

predict the rotation curve of a disk galaxy given measurements of the density of the disk. That prediction is easily testable using the observed motions of stars or gas clouds in the disk, and this is true for any of the (hundreds or thousands) of well-observed disk galaxies.

The standard-model hypothesis SCM1 makes no prediction about the behavior of *observable* matter. It only says something about the *dark* matter. An observed rotation curve can not be used to test hypothesis SCM1; rather, that hypothesis instructs the scientist to *use* the measured $V(R)$ (together with the observed density of stars and gas in the disk) to predict the dark matter distribution. The rotation curve is treated as part of the background knowledge. In Milgrom's (1989, p. 216) words, postulate SCM1 "simply states that dark matter is present in whatever quantities and space distribution is needed to explain away whichever mass discrepancy arises." Worrall's rule: "One can't use the same fact twice: once in the construction of a theory and then again in its support" tells us that the rotation curve, by virtue of having been used in the construction of the dark matter distribution, has lost its ability to provide support to the dark matter hypothesis.

This lack of testability extends to cases in which a galaxy's rotation curve is supplemented by other kinds of kinematical data. One much-discussed example is the so-called 'Oort problem': understanding what the observed distribution of stellar velocities *perpendicular* to the Milky Way disk (at the Solar circle) implies about the local mass density in the disk. Milgrom's theory makes a solid prediction (this is prediction no. 4 in the list of the previous section) and that prediction has been confirmed: that is, the vertical motions are observed to be consistent, under the modified dynamics, with the observed ('baryonic') mass in the disk (Nipoti et al. 2007; Bienaymé et al. 2009).

Under the standard model, one could imagine using SCM1, together with the observed Milky Way rotation curve, to estimate the dark matter density near the Sun; then test whether the vertical force generated by the total local density ('baryonic' plus dark) correctly predicts the observed vertical motions. But standard-model cosmologists have never succeeded in doing this. In the first such studies, Bahcall (1984a, b) assumed a spherical dark 'halo' having a mass distribution designed to explain the Galaxy's rotation curve. That model made a definite prediction about the local dark matter density, as well as its dependence on distance above or below the disk. But Bahcall recognized the degeneracy of the dark matter models:

> Since we haven't yet observed the unseen material, we don't know how it is distributed. Therefore we have to try different models for the unseen material to see how the results depend upon our assumptions (Bahcall 1987, p. 19).

And indeed he treated the local dark matter density and its dependence on z as adjustable quantities.[10] Subsequent studies of the Oort problem by standard-model cosmologists have likewise shied away from casting their analyses as tests. For instance, Smith et al. (2012) write (italics added):

---

[10] Although Bahcall never claimed to be testing a standard-model prediction, he did note (Bahcall 1987) that the data were explainable via Milgrom's theory.

> If we assume our background mass represents the dark halo, it corresponds to a local dark matter density of 0.57 GeV cm^{-3}, which is noticeably larger than the canonical value of 0.30 GeV cm^{-3} typically assumed …As pointed out by various authors …, *the local dark matter density is uncertain by a factor of at least two*. Our analysis adds still more weight to the argument that the local halo density may be substantially underestimated by the canonical value of 0.30 GeV cm^{-3} (Smith et al. 2012, p. 11).[11]

The local value of ρ_{DM}—the prediction that could be refuted via an analysis of the vertical motions – is typically decoupled from the rotation curve constraint in these studies by allowing the dark matter halo to be nonspherical (e.g. Garbari et al. 2012; Bienaymé et al. 2014). By treating the halo axis ratio as an extra, freely adjustable parameter, many values of the local dark matter density can be made consistent with a given rotation curve, thus effectively nullifying any predictive power of SCM1. (This degeneracy adds to the "factor of at least two" uncertainty mentioned in the quotation from Smith et al.) And indeed some studies are forced to assume extremely contrived shapes for the dark matter halo in order to get the vertical kinematics 'right' (Read 2014).

One way to make a *testable* prediction of the standard-model postulate is to couple SCM1 with some other hypothesis. In fact, standard-model cosmologists routinely assume, in addition to SCM1, that

SCM2 The dark matter of SCM1 is composed of elementary particles.

(E.g. Funk (2015, p. 12264): "Today, it is widely accepted that dark matter exists and that it is very likely composed of elementary particles, which are weakly interacting and massive.") One novel prediction immediately follows: some of the dark particles associated with the Milky Way must be passing at every moment through an Earth-based laboratory and could be detected, through their interaction with normal matter. But this prediction—while capable of being confirmed—is not refutable, since nothing whatsoever is known about the properties of the putative particles (aside from the fact that no *known* particles have properties that would make them acceptable candidates). A failure to detect the particles might simply mean that their cross-section for interaction with the normal matter in the detectors is very small, and that is in fact one explanation that particle physicists propose for their almost four-decade failure to detect a signal (Bertone and Hooper 2018).[12]

In this respect, the particle dark matter hypothesis is in a state similar to that of the atomistic hypothesis at the end of the 19th century. Popper (1983, p. 191) noted that the hypothesis that atoms exist was, for a long time, too vague to be refuted:

---

[11] 0.30 GeV $\mathrm{cm}^{-3} \approx 0.008 M_{\odot}\ \mathrm{pc}^{-3}$.

[12] When Elena Aprile was asked to estimate a cost for her XENONnT dark matter experiment at the Italian Gran Sasso National Laboratory, the *New York Times* reports that she "was reluctant to put a price on the project. An earlier version of the experiment with 3.3 tons of xenon cost $30 million. But that didn't include the people, she said. A big part of the cost is xenon itself, which costs around $2 million per ton, she added. Her new detector will have 8.5 tons" (Overbye 2020). There are about a half-dozen such experiments currently underway (as reviewed by Kisslinger and Das 2019). Given that the hypothesis being tested by the direct-detection experiments (that dark particles are passing through the laboratory) is not refutable, it is reasonable to ask what will have been accomplished by those experiments assuming the continued absence of a detection.

> Failure to detect the corpuscles, or any evidence of them, could always be explained by pointing out that they were too small to be detected. Only with a theory that led to an estimate of the size of the molecules was this line of escape more or less blocked, so that refutation became in principle possible.

Popper's statement is perfectly applicable to the (particle) dark matter hypothesis if one replaces 'size of the molecules' by 'cross section of interaction of the dark particles with normal matter.'

We are now in a position to assess how well postulates SCM1 and SCM2 accord with Popper's methodological rules. Rules nos. 7, 11 and 19 are violated: as we have seen, the postulates have little if any testable, that is, refutable, content. Rule no. 25, which demands that the modified theory "should pass some new, and severe, tests," is not (yet) satisfied. And I would argue that rules no. 8 ("*criteria of refutation* have to be laid down beforehand) and no. 9 ("in the case of a threat to our system, we will not save it by any kind of *conventionalist stratagem*") are also violated. Indeed a pervasive feature of the standard-model literature is the conviction that any anomaly will, eventually, be explainable within the paradigm.[13] This attitude often conflicts with rule no. 13 as well, which forbids the appeal to "derivations to be discovered in the future."

It is fair to say that rules no. 14 and15 *are* satisfied since the number of additional axioms (two in this case) is small.[14] But I would insist that rule no. 16 ("should proceed from some simple, new, and powerful" idea) is violated. Here is one way to justify that statement: When teaching introductory astrophysics, a problem that is commonly set to students (who have not yet learned about dark matter) is to ask them what can be inferred from the asymptotic flatness of rotation curves. The 'correct' answer, the answer that students are expected to find, is: there must be non-luminous matter in or around the galactic disk. Far from being a "new, and powerful" idea, dark matter is almost literally the first explanation that pops into anyone's head.

Finally, rule no. 26 recommends, when deciding between two theories, to choose the theory that is "testable in the most rigorous way." That recommendation would clearly favor Milgrom's theory over the standard model, at least in terms of their postulates that target the rotation curve anomaly.

---

[13] A striking example is the standard-model response to the remarkably correlated distribution of satellite galaxies around the Milky Way and the Andromeda galaxy, observations that have no, even remotely, plausible explanation under that model. Kroupa (2016, p. 557) documents the variety of 'conventionalist stratagems' adopted by standard-model cosmologists in response to those observations and concludes, "The [standard-model] community appears to have developed an unhealthy sense of simply ignoring or burying previously obtained results if these are highly inconsistent with the standard model of cosmology." While MOND does not make a clear prediction here, the observed correlations do not constitute a prima facie problem for Milgrom's theory (Pawlowski 2018).

[14] On the other hand, one could reasonably take the point of view that postulate SCM1 comprises a very *large* number of independent postulates, since the specification of the dark matter distribution around any single galaxy requires a 3d function, and furthermore a function that is different for every galaxy.

6 .

Rule no. 7 (Popper 1959, p. 56) states

> only such statements may be introduced in science as are inter-subjectively testable

and rule no. 24 (Popper 1959, p. 268) is

> we shall not continue to accord a positive degree of corroboration to a theory which has been falsified by an inter-subjectively testable experiment.

Of course, as Popper acknowledged, and as others (Kuhn, Lakatos, Feryerabend) also emphasized, a falsifying instance—even an inter-subjectively accepted one—need not signal the ultimate death of a theory; indeed many of the rules in Table 1 are guidelines for the scientist seeking to *modify* her theory in response to a falsification.

But a theory that is not testable is not falsifiable. Nevertheless, standard-model cosmologists do acknowledge that their theory is inconsistent with a number of well-established facts. Silk and Mamon (2012) list seventeen such inconsistencies; Bullock and Boylan-Kolchin (2017) list about a dozen; and Kroupa (2012) gives twenty-two. Which, if any, of these instances constitute falsifications in the sense that Popper used that term?

I would argue that there have been only two important instances (since the 1960s) where the standard cosmological model has made predictions that were inter-subjectively testable; and that in both cases, the predictions were subsequently contradicted by observations.

The first instance pre-dated the dark matter postulates: it was the demonstration that galaxy rotation curves are not correctly predicted by Newton's theory (Rubin et al. 1978; Bosma 1981).

The second occurred twenty years later: the discovery that the cosmological expansion is accelerating rather than decelerating, as Einstein's equations generically predict (Riess et al. 1998; Perlmutter et al. 1999).[15]

One feature that sets these two failures of prediction apart is the way the standard-model community chose to respond to them. In both cases, a (nearly) immediate and (nearly) unanimous consensus was reached that an auxiliary hypothesis should be added to the theory: 'dark matter' (that is, SCM1) in the first case, 'dark energy' in the second (Longair 2006). These two hypotheses were designed to maintain the integrity of Einstein's (or Newton's) theory of gravity in the face of falsifying data, and indeed the assumed properties of both dark matter and dark energy have been revised a number of times, as needed to maintain that integrity as new data emerged (Wang 2010; Majumdar 2015). Following Lakatos (1978), we can therefore identify Einstein's (Newton's) theory of gravity as constituting part of the 'hard core' of the standard cosmological 'research program,' and the postulates relating to dark matter and dark energy as part of the 'protective belt' of auxiliary hypotheses that serve to maintain the integrity of that hard core in the face of refutations.

---

[15] Interestingly, both of these instances can be seen as failed tests of Einstein's theory in the low-acceleration regime.

By contrast, almost all of the other standard-model problems listed by Silk & Mamon, Bullock & Boylan-Kolchin and Kroupa are failures of *accommodation*, not of *prediction*.

What I mean by "accommodation" is best illustrated by an example. Consider disk galaxy rotation curves. As noted above, dark matter postulate SCM1 makes no prediction about rotation curves, nor have standard-model cosmologists yet come up with any scheme that allows them to predict the rotation curve of any galaxy. However, standard-model theorists have devoted enormous effort to 'getting rotation curves right': that is: to finding ways to simulate the formation and evolution of galaxies starting from early times, such that the *statistical* properties of the *simulated* galaxies match (in some specified sense) the statistical properties of real galaxies—"properties" here defined, of course, in terms of the observable matter (stars, gas). When a standard-model cosmologist says that he has failed to predict the rotation curves of (say) dwarf galaxies, what he means is that he has not been able to find a physically reasonable set of simulation parameters that results in simulated galaxies whose structure and kinematics 'look', in some average sense, like those of observed dwarf galaxies.

From a Milgromian perspective, these standard-model failures reflect the difficulty of getting the dark matter in the simulations to behave 'correctly'—to distribute itself around every galaxy so that its stars and gas respond to the total gravitational force in a manner that mimics the modified dynamics. Standard-model theorists, by contrast, consider the behavior of the dark matter in their simulations to be unproblematic; the problem, as seen by them, is to get those pesky *baryons* to behave. (E. g. Bullock and Boylan-Kolchin (2017, p. 380): "Within the standard ΛCDM model, most properties of small-scale structure can be modeled with high precision in the limit that baryonic physics is unimportant." "Small-scale" here refers to single galaxies.)

Nowhere are the standard-model failures of accommodation more striking than in the case of dwarf galaxies, which (they would say) are 'dark-matter dominated': that is: fully in the Milgromian regime. And it is the dwarf galaxy literature that provides some of the starkest illustrations of just how far standard-model cosmologists have strayed from methodological rule no. 7, that "only such statements may be introduced in science as are inter-subjectively testable":

Standard-model cosmologists identify the retinue of observed dwarf galaxies orbiting the Milky Way with the dark matter 'sub-halos' that form in their simulations. A problem immediately arises: the number of sub-halos in the simulations far exceeds the number of observed satellite galaxies (e.g. Silk and Mamon (2012, 939): "The excessive predicted numbers of dwarf galaxies are [sic] one of the most cited problems with ΛCDM. The discrepancy amounts to two orders of magnitude.") This discrepancy is called by standard-model cosmologists the 'missing-satellites problem,'[16] and most attempts to solve it invoke some mechanism for heating or removing gas from the sub-halos before the epoch of star formation. No single mech-

---

[16] Terminology like this should be of interest to social epistemologists: it suggests that standard-model cosmologists, when conceptualizing the physical world, privilege their simulations over the actual data. The name that *Milgromian* researchers attach to this standard-model failure is the 'dwarf over-prediction problem.' Milgromian researchers postulate a different origin for the satellite

anism ‘works’ across the full spectrum of dwarf galaxy (that is, sub-halo) masses, and standard-model cosmologists blithely invoke different mechanisms, as the need arises, to explain the data on different mass scales. For instance, Bullock (2010, p. 12), after listing the various mechanisms that have been proposed for suppressing star formation in the sub-halos, remarks:

> each imposes a different mass scale of relevance …If, for example, we found evidence for very low-mass dwarf galaxies $V_{\max} \sim 5$ km s^{-1} then these [galaxies] would be excellent candidates for primordial H_2 cooling ‘fossils’ of reionization in the halo.

And Bullock and Boylan-Kolchin (2017, 370) write that “while many independent groups are now obtaining similar results in cosmological simulations of dwarf galaxies . . . this is not an ab initio ΛCDM prediction, and it depends on various adopted parameters in galaxy formation modeling.”

As Karl Popper (1983, p. 168) remarked in his critique of Freud’s theory, “*every conceivable case will become a verifying instance*” (italics his).

7 .

Rules nos. 11, 18, 19, 25, 26 and 29 direct the scientist to prefer theories with greater explanatory power, content, or testability. Popper considered these qualities to be closely linked, and in *Conjectures and Refutations* (p. 217) he called the requirement that theories evolve in the direction of increasing content the criterion of “potential satisfactoriness” or “potential progress”. He defined a criterion of *actual* scientific progress in terms of rules nos. 25 and 26: that is: the requirement that at least some of a theory’s new content be experimentally confirmed (p. 220). As is well known, Lakatos (1978) followed Popper’s lead in defining the “empirical progressivity” of an evolving theory in terms of corroborated excess content—that is: confirmed, novel predictions.

Popper (1962) sought to strengthen his intuitive idea of progress by defining the “verisimilitude” or “truthlikeness” of a theory, a measure of the theory’s closeness to truth. Given some definition of verisimilitude, and some scheme for evaluating it, progress could be identified with an increase in a theory’s verisimilitude. But Popper did not view verisimilitude as a concept that should necessarily take the place of, or transcend, methodological considerations:

> I do not suggest that the explicit introduction of the idea of verisimilitude will lead to any changes in the theory of method. On the contrary, I think that my theory of testability or corroboration by empirical tests is the proper methodological counterpart to this new metalogical idea. The only improvement is one of clarification (Popper 1963, p. 235).

(In Watkins’s (1978, p. 365) words: “Popper got along well enough without the idea of verisimilitude for a quarter century after 1934.”) Nevertheless Popper’s “new met-

---

galaxies—see Kroupa (2012) – and the small number of satellites observed around the Milky Way constitutes no problem for them.

alogical idea" has been enthusiastically taken up by a generation of realist philosophers, including Oddie (1986), Niiniluoto (1987), Kieseppä (1996), Kuipers (2000), Zwart (2011) and others.

In view of this impressive body of work, it is natural to ask whether it is now feasible to compare the two theories of cosmology in terms of their verisimilitude. I will argue that the answer is 'no' and furthermore that the prospects for doing so in the future are bleak.

My first point has to do with the way that scientific theories (including Milgrom's) typically evolve. As Feyerabend (2010, p. 157) noticed,

> Theories which effect the overthrow of a comprehensive and well-entrenched point of view. . . are initially restricted to a fairly narrow domain of facts, to a series of paradigmatic phenomena which lend them support, and they are only slowly extended to other areas.

Milgrom's theory in its current state is marvelously successful at making novel predictions for galaxies and groups of galaxies, and it has also had some notable successes in anticipating large-scale data (Merritt 2020, Chap. 6). But there is general agreement even among Milgromian theorists that a suitable, general relativistic version of the theory (or its equivalent) is not yet available; in the language of Imre Lakatos, the Milgromian research program is in an earlier stage of development than the standard cosmological research program.[17] And so comparisons with standard-model explanations of large-scale structure or the high-redshift universe would be pointless and, in all likelihood, misleading.

One might hope to sidestep this difficulty by comparing the verisimilitude of the two theories only in some narrow regime where both claim to make predictions; for instance, the internal kinematics of dwarf galaxies (e.g. Lazutkina (2017) who applies Niiniluoto's (1999) measure of truthlikeness to velocity data for a set of dwarfs). But such a project runs solidly up against an intractable problem.[18] Milgrom's theory is quite capable of making inter-subjectively testable predictions about galaxies. The standard cosmological model—due largely to the vagueness of the dark matter hypothesis – is not; as we have seen, the best it can hope for is to include a scheme for simulating galaxy evolution that leads to model galaxies that look, in some average or statistical sense, like real galaxies. But even standard-model cosmologists can be quite frank about the degree to which their numerical experiments are *explicitly designed to reproduce known facts*. For instance, one researcher writes[19]

> Galaxy formation simulations …tune parameters such that the simulations produce realistic-looking galaxy populations. In this sense the sub-grid models are 'validated' as 'realistic'

---

[17] This difference reflects the enormous disparity in number of scientists working in the two research programs, as well as the disinclination of government agencies to fund Milgromian researchers, among other possible factors.

[18] The difficulty discussed in this paragraph exists for any criterion of success that is essentially empirical or instrumentalist, e.g. Carnap's (1950) 'qualified instance confirmation,' Laudan's (1978) 'problem-solving efficiency,' van Fraassen's (1980) 'constructive empiricism' etc.

[19] Quoted by Merritt (2020, p. 75) who gives the source. "Sub-grid models" refers to algorithms that are meant to represent, in some approximate manner, physical processes that occur on scales of time or space that are far too small to be simulated directly, e.g. turbulence, stellar winds etc.

> models by plausibility arguments in comparison to observations. Historically these models result from trial and error experiments. The models themselves might easily be ‘wrong’ (in a strict physical sense) or assuming unrealistically high values for the coupling efficiencies—they still produce realistic galaxy properties and the authors claim success.

The ‘success’ of standard-model cosmologists at explaining observations of galaxies is a function of a host of factors that are external to their theory: how creative they were in crafting the “sub-grid models”; how effectively they were able to convince the larger community (including, most importantly, journal referees and editors) of the physicality of those models; how much time (human and computer) was available for the simulations and their analysis (and, therefore, how much funding was available); etc. As Thomas Kuhn might have said, these are tests of the theorist, not of the theory.

Given the fundamentally different ways in which the two groups of cosmologists achieve correspondence of their theory with the facts, it is reasonable to ask whether case studies of methodology might not be a better guide to the progress-toward-truth of their respective theories than measures of verisimilitude. After all, one need not have a perfect criterion of justice (say) to know that there are certain methodologies (e.g. deposition of witnesses) that are more conducive to *achieving* justice than others (e.g. divination). In the same way, it is hard to believe that a critical, or falsificationist, approach to theory testing is less likely to lead to true theories than an uncritical, or verificationist, approach (e.g. Agassi 1959; Popper 1962; Albert 1987; Gadenne 2006).

8 .

Niiniluoto (1999, p. 17) speculates about why a scientist would choose to follow a methodology like Popper’s:

> …for centuries, theory and practice have already been in a mutual interaction in the field of scientific inference. Scientists learn to do science through implicit indoctrination and explicit instruction from their masters, textbooks, and colleagues. So if a case study reveals that a group of real scientists favours ‘bold hypotheses’ and ‘severe tests’, we may judge that they, or their teachers, have read Popper.

I am quite certain that Niiniluoto is mistaken here. First of all, he is crediting philosophers with far too much influence. Most scientists—particularly young scientists, but also the scientists who write the textbooks—are dismissive of philosophy, not to say contemptuous of it. Scientists have reasons for doing the things they do, of course, but they don’t get those reasons from the philosophers.

More to the point: Niiniluoto’s explanation would imply that only *Milgromian* researchers have been brought up as critical rationalists, while the bulk of cosmologists have been “indoctrinated” into some other epistemological school (Niiniluoto’s inductivism, perhaps). And *that* hypothesis is easily debunked: The number of Milgromian researchers is quite small (perhaps two dozen worldwide, certainly not many more); I know most of them personally; and I can attest that their educations were quite of a piece with the educations of the standard-model cosmologists in their cohorts. There exists no secret society that is indoctrinating selected young scientists into the Popperian mysteries.[20]

Here is what does impress a scientist: a bold new conjecture that bears fruit. The paradigmatic example, one that every physical scientist learns about early in their education, is the set of postulates from which Bohr derived the energy levels of the hydrogen atom. Bohr's success is impressive because it was so improbable. Einstein (speaking at a time when Popper was eleven years old) declared that "There must be something behind it. I do not believe that the derivation of the absolute value of the Rydberg constant is purely fortuitous."[21] And it is obvious to any beginning student of quantum mechanics that a 'turn-the-crank' methodology like abduction or inference-to-the-best-explanation could not possibly have led Bohr to his bold conjecture, a conjecture that went far beyond the evidence that motivated it.

I am sure that standard-model cosmologists are just as impressed as other scientists by instances in which a bold hypothesis survives a severe test. But the standard cosmological model (at least since the addition of dark matter ca. 1980) is simply not suited to making testable predictions, much less bold ones. So standard-model cosmologists have, understandably, resigned themselves to the post hoc accommodation of new data, typically via large-scale computer simulations, and typically only in a statistical sense. Whereas Milgrom's bold theory *is* eminently testable, even using data from a single galaxy, and (as we have seen) its novel predictions have again and again survived attempts to refute them. One need look no farther to understand why Milgromian researchers have stuck with a methodology that aligns with Popper's.

References

Agassi, J.: Epistemology as an aid to science: comments on Dr. Buchdal's paper. British J. Phil. Sci. **X**(38), 135–146 (1959)

Albert, H.: Science and the search for truth: critical rationalism and the methodology of science. In: Jarvie, I.C., Charles, I., Agassi, J. (eds.) Rationality: The Critical View, pp. 69–82. Nijhoff, Dordrecht (1987)

Bahcall, J.N.: K giants and the total amount of matter near the sun. Astrophys. J. **287**, 926–944 (1984a)

Bahcall, J.N.: Self-consistent determinations of the total amount of matter near the sun. Astrophys. J. **276**, 169–181 (1984b)

Bahcall, J.N.: Dark matter in the galactic disk. In: Kormendy, J., Knapp, G.R. (eds.) Dark Matter in the Universe. IAU Symposium, vol. 117, pp. 17–27 (1987)

Begeman, K.G., Broeils, A.H., Sanders, R.H.: Extended rotation curves of spiral galaxies-dark haloes and modified dynamics. Monthly Notices R. Astron. Soc. **249**, 523–537 (1991)

Bertone, G., Hooper, D.: History of dark matter. Rev. Mod. Phys. **90** (2018)

---

[20] Niiniluoto's "implicit indoctrination" calls to mind Tolstoy's (1906) invocation of "epidemic suggestions" to explain the (to him) unfathomable popularity of Shakespeare's plays.

[21] Quoted by Jammer (1966, p. 86). Jammer gives the original German in his note 107 as "da muß etwas dahinter sein; ich glaube nicht, daß die Rydbergkonstante durch Zufall in absoluten Werten ausgedrückt richtig herauskommt.".

Bienaymé, O., et al.: Weighing the local dark matter with RAVE red clump stars. Astron. Astrophys. **571**, 92 (2014)

Bienaymé, O., Famaey, B., Wu, X., Zhao, H.S., Aubert, D.: Galactic kinematics with modified Newtonian dynamics. Astron. Astrophys. **500**, 801–805 (2009)

Bosma, A.: 21-cm line studies of spiral galaxies. I. Observations of the galaxies NGC 5033, 3198, 5055, 2841, and 7331. Astron. J. **86**, 1791–1824 (1981)

Bullock, J.S.: Notes on the missing satellites problem. arXiv e-prints, 1009-4505 (2010)

Bullock, J.S., Boylan-Kolchin, M.: Small-scale challenges to the ΛCDM paradigm. Ann. Rev. Astron. Astrophys. **55**, 343–387 (2017)

Carnap, R.: Logical Foundations of Probability. Routledge and Kegan Paul, London (1950)

de Bernardis, P., et al.: Multiple peaks in the angular power spectrum of the cosmic microwave background: significance and consequences for cosmology. Astrophys. J. **564**, 559–566 (2002)

de Blok, W.J.G., McGaugh, S.S.: The dark and visible matter content of low surface brightness disc galaxies. Monthly Notices R. Astron. Soc. **290**, 533–552 (1997)

Donato, F., Gentile, G., Salucci, P., Frigerio Martins, C., Wilkinson, M.I., Gilmore, G., Grebel, E.K., Koch, A., Wyse, R.: A constant dark matter halo surface density in galaxies. Monthly Notices R. Astron. Soc. **397**, 1169–1176 (2009)

Feyerabend, P.: Against Method, 4th edn. Verso, London (2010)

Fields, B.D.: The primordial lithium problem. Ann. Rev. Nucl. Part. Sci. **61**, 47–68 (2011)

Funk, S.: Indirect detection of dark matter with γ rays. Proc. Natl. Acad. Sci. **112**, 12264–12271 (2015)

Gadenne, V.: Methodological rules, rationality, and truth. In: Cheyne, C., Worrall, J. (eds.) Rationality and Reality: Conversations with Alan Musgrave, pp. 97–107. Springer, Dordrecht (2006)

Garbari, S., Liu, C., Read, J.I., Lake, G.: A new determination of the local dark matter density from the kinematics of K dwarfs. Monthly Notices R. Astron. Soc. **425**, 1445–1458 (2012)

Hempel, C.G.: Studies in the logic of confirmation. Mind **54**, 1–2697121 (1945)

Hosiasson-Lindenbaum, J.: On confirmation. J. Symbolic Logic **5**, 133–148 (1940)

Jaffe, A.H., et al.: Cosmology from MAXIMA-1, BOOMERANG, and COBE DMR cosmic microwave background observations. Phys. Rev. Lett. **86**, 3475–3479 (2001)

Jammer, M.: The Conceptual Development of Quantum Mechanics. McGraw-Hill, New York (1966)

Jarvie, I.: The Republic of Science: The Emergence of Popper's Social View of Science 1935–1945. Rodopi, Amsterdam (2001)

Johansson, I.: A Critique of Karl Popper's Methodology. Akademif orlaget, Stockholm (1975)

Keuth, H.: The Philosophy of Karl Popper. Cambridge University Press, Cambridge (2005)

Kieseppä, I.A.: Truthlikeness for Multidimensional. Quantitative Cognitive Problems. Kluwer, Dordrecht (1996)

Kisslinger, L.S., Das, D.: A brief review of dark matter. Int. J. Mod. Phys. A **34**, 1930013 (2019)

Ko, C.-M.: Dark matter. In: Ko, C.-M., Yu, P.-C., Chang, C.-K. (eds.) Astronomical Society of the Pacific Conference Series, vol. 513, p. 217 (2018)

Kroupa, P.: The dark matter crisis: falsification of the current standard model of cosmology. Publ. Astron. Soc. Australia **29**, 395–433 (2012)

Kroupa, P.: In: D'Onofrio, M., Rampazzo, R., Zaggia, S. (eds.) From the Realm of the Nebulae to Populations of Galaxies, pp. 547–566. Springer, Berlin (2016)

Kuhn, T.S.: The Structure of Scientific Revolutions. The University of Chicago Press, Chicago (1962)

Kuipers, T.A.F.: From Instrumentalism to Constructive Realism: On Some Relations between Confirmation, Empirical Progress, and Truth Approximation. Kluwer, Dordrecht (2000)

Lakatos, I.: Changes in the problem of inductive logic. In: Lakatos, I. (ed) The Problem of Inductive Logic. Proceedings of the International Colloquium in the Philosophy of Science, vol. 2, pp. 315–417 (1968)

Lakatos, I.: Falsification and the methodology of scientific research programmes (1970). Lakatos **8–101** (1978)

Lakatos, I.: Introduction: science and pseudoscience (1973). Lakatos **1–7** (1978)

Lakatos, I.: The methodology of scientific research programmes (philosophical papers, vol. I). Worrall, J., Currie, G. (eds). Cambridge University Press, Cambridge (1978)

Laudan, L.: Progress and its Problems: Towards a Theory of Scientific Growth. University of California Press, Berkeley (1978)

Lazutkina, A.: Theoretical Terms and Theoretical Objects of Contemporary Astrophysics: A Logico-Methodological Analysis. Bachelor's thesis, Lomonosov Moscow State University (2017)

Leibniz, G.W.: Letter to Herman Conring, 19 March. In: Loemker, L. (ed) Gottfried Wilhelm Leibniz: Philosophical Papers and Letters, 2nd edn, pp. 186–191. Reidel, Dordrecht (1678) (1970)

Lelli, F., McGaugh, S.S., Schombert, J.M.: The small scatter of the baryonic Tully-Fisher relation. Astrophys. J. Lett. **816**, 14 (2016a)

Lelli, F., McGaugh, S.S., Schombert, J.M., Pawlowski, M.S.: The relation between stellar and dynamical surface densities in the central regions of disk galaxies. Astrophys. J. Lett. **827**, 19 (2016b)

Lelli, F., McGaugh, S.S., Schombert, J.M., Pawlowski, M.S.: One law to rule them all: the radial acceleration relation of galaxies. Astrophys. J. **836**, 152 (2017)

Liu, J., Chen, X., Ji, X.: Current status of direct dark matter detection experiments. Nat. Phys. **13**(3), 212–216 (2017)

Longair, M.S.: The Cosmic Century: A History of Astrophysics and Cosmology. Cambridge University Press, Cambridge (2006)

Losee, J.: Theories of Scientific Progress. Routledge, New York (2004)

Losee, J.: Theories on the Scrap Heap. University of Pittsburgh, Pittsburgh (2005)

Majumdar, D.: Dark Matter. CRC Press, Boca Raton (2015)

McGaugh, S.S.: Distinguishing between cold dark matter and modified Newtonian dynamics: predictions for the microwave background. Astrophys. J. Lett. **523**, 99–102 (1999)

McGaugh, S.S.: Milky way mass models and MOND. Astrophys. J. **683**, 137–148 (2008)

McGaugh, S.S.: Novel test of modified Newtonian dynamics with gas rich galaxies. Phys. Rev. Lett. **106** (2011)

McGaugh, S.S.: A tale of two paradigms: the mutual incommensurability of λcdm and mond. Can. J. Phys. **93**, 250–259 (2015)

McGaugh, S.S., Lelli, F., Schombert, J.M.: Radial acceleration relation in rotationally supported galaxies. Phys. Rev. Lett. **117** (2016)

Merritt, D.: Cosmology and convention. Stud. History Phil. Mod. Phys. **57**, 41–52 (2017)

Merritt, D.: A Philosophical Approach to MOND: Assessing the Milgromian Research Program in Cosmology. Cambridge University Press, Cambridge (2020)

Milgrom, M.: A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. Astrophys. J. **270**, 365–370 (1983a)

Milgrom, M.: A modification of the Newtonian dynamics: implications for galaxies. Astrophys. J. **270**, 371–383 (1983b)

Milgrom, M.: A modification of the Newtonian dynamics: implications for galaxy systems. Astrophys. J. **270**, 384–389 (1983c)

Milgrom, M.: On the use of galaxy rotation curves to test the modified dynamics. Astrophys. J. **333**, 689 (1988)

Milgrom, M.: Alternatives to dark matter. Comments Astrophys. **13**, 1–2 (1989)

Milgrom, M.: MOND impact on and of the recently updated mass-discrepancy-acceleration relation. arXiv e-prints, 1609-06642 (2016)

Miller, D.: Critical Rationalism: A Restatement and Defence. Open Court, Chicago (1994)

Netterfield, C.B., et al.: A measurement by BOOMERANG of multiple peaks in the angular power spectrum of the cosmic microwave background. Astrophys. J. **571**, 604–614 (2002)

Niiniluoto, I.: Truthlikeness. Reidel, Dordrecht (1987)

Niiniluoto, I.: Critical Scientific Realism. Oxford University Press, New York (1999)

Niiniluoto, I.: Truth-Seeking by Abduction. Springer, Cham (2018)

Nipoti, C., Londrillo, P., Zhao, H., Ciotti, L.: Vertical dynamics of disc galaxies in modified Newtonian dynamics. Monthly Notices R. Astron. Soc. **379**, 597–604 (2007)

Oddie, G.: Likeness to Truth. Reidel, Dordrecht (1986)

Overbye, D.: Cosmic Secret May Have To Wait. New York Times, April 7, 2020

Pawlowski, M.S.: The planes of satellite galaxies problem, suggested solutions, and open questions. Mod. Phys. Lett. A **33**, 1830004 (2018)

Perlmutter, S., et al.: Measurements of Ω and Λ from 42 high-redshift supernovae. Astrophys. J. **517**, 565–586 (1999)

Popper, K.: Logik der Forschung. Springer, Vienna (1934)

Popper, K.: The Open Society and Its Enemies, vol. 1. The Spell of Plato, 4th edn. Routledge, London (1945)

Popper, K.: The Poverty of Historicism. Routledge & Kegan Paul, London (1957)

Popper, K.: The Logic of Scientific Discovery. Basic Books, New York (1959)

Popper, K.: Some comments on truth and the growth of knowledge. In: Nagel, E., Suppes, P., Tarski, A. (eds.) Logic, Methodology and Philosophy of Science, pp. 285–292. Stanford University Press, Stanford (1962)

Popper, K.: Conjectures and Refutations: The Growth of Scientific Knowledge. Routledge & Kegan Paul, London (1963)

Popper, K.: Objective Knowledge: An Evolutionary Approach. Oxford University Press, Oxford (1972)

Popper, K.: Realism and the Aim of Science. Rowman and Littlefield, Totowa, NJ (1983)

Psillos, S.: Scientific Realism: How Science Tracks Truth. Routledge, London (1999)

Read, J.I.: The local dark matter density. J. Phys. G Nucl. Phys. **41** (2014)

Riess, A.G., et al.: Observational evidence from supernovae for an accelerating universe and a cosmological constant. Astron. J. **116**, 1009–1038 (1998)

Rubin, V.C., Ford, W.K., Thonnard, N.: Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa → Sc. Astrophys. J. Lett. **225**, 107–111 (1978)

Schneider, P.: Extragalactic Astronomy and Cosmology, 2nd edn. Springer, Berlin (2015)

Shull, J.M., Smith, B.D., Danforth, C.W.: The baryon census in a multiphase intergalactic medium: 30% of the baryons may still be missing. Astrophys. J. **759**, 23 (2012)

Silk, J., Mamon, G.A.: The current status of galaxy formation. Res. Astron. Astrophys. **12**, 917–946 (2012)

Smith, M.C., Whiteoak, S.H., Evans, N.W.: Slicing and dicing the Milky Way disk in the Sloan digital sky survey. Astrophys. J. **746**, 181 (2012)

Spergel, D.N., et al.: Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: implications for cosmology. Astrophys. J. Suppl. **170**, 377–408 (2007)

Stark, D.V., McGaugh, S.S., Swaters, R.A.: A first attempt to calibrate the baryonic Tully-Fisher relation with gas-dominated galaxies. Astron. J. **138**, 392–401 (2009)

Tolstoy, L.: Shakespeare and the drama (Part I). Fortnightly Rev. **80**(480), 964–983 (1906)

Trachternach, C., de Blok, W.J.G., McGaugh, S.S., van der Hulst, J.M., Dettmar, R.-J.: The baryonic Tully-Fisher relation and its implication for dark matter halos. Astron. Astrophys. **505**, 577–587 (2009)

van Fraassen, B.C.: The Scientific Image. Clarendon Press, Oxford (1980)

Wang, Y.: Dark Energy. Wiley-VCH, Weinheim (2010)

Watkins, J.: Corroboration and the problem of content-comparison. In: Radnitzky, G., Andersson, G. (eds.) Boston Studies in the Philosophy of Science, vol. 58. Progress and Rationality in Science, pp. 339–378. Reidel, Dordrecht (1978)

Worrall, J.: The ways. In: Radnitzky, G., Andersson, G. (eds.) Boston Studies in the Philosophy of Science, vol. 58, Which the Methodology of Scientific Research Programmes Improves on Popper's Methodology, in Progress and Rationality in Science, pp. 45–70. Reidel, Dordrecht (1978)

Worrall, J.: Scientific discovery and theory-confirmation. In: Pitt, J.C. (ed.) Change and Progress in Modern Science, pp. 301–331. Reidel, Dordrecht (1985)

Wu, X., Kroupa, P.: Galactic rotation curves, the baryon-to-dark-halo-mass relation and space-time scale invariance. Monthly Notices R. Astron. Soc. **446**, 330–344 (2015)

Zahar, E.: Why did Einstein's research programme supersede Lorentz's? British J. Phil. Sci. **24**, 95–123223262 (1973)
Zwart, S.D.: Refined Verisimilitude. Springer, Dordrecht (2011)

David Merritt was a professor at the Rochester Institute of Technology, USA and before that at Rutgers University, whose research interests include galaxy dynamics and evolution, supermassive black holes, and computational astrophysics. He is a former Chair of the Division on Dynamical Astronomy of the American Astronomical Society and is a founding member of the Center for Computational Relativity and Gravitation at RIT. He is author of *Dynamics and Evolution of Galactic Nuclei* (PUP 2017) and *A Philosophical Approach to MOND: Assessing the Milgromian Research Program in Cosmology* (CUP 2020).

The Application of Popperian Methodology to Contemporary Cosmology

Anastasiia Lazutkina

1 Introduction

The scientific methodology developed by Karl Popper has been highly influential not only among philosophers of science but among practicing scientists themselves. Contemporary cosmology is not an exception. As Helge Kragh notes in his contribution to this volume, prominent cosmologists and other physicists have appealed to Popper's falsifiability criterion in an effort to combat what they consider to be unscientific approaches to doing physics.[1] Others have expressed disapproval of the idea that rigid rules devised by philosophers could restrict the activity of a scientific research community.

Ironically, none of the cosmologists appealing to Popper's methodological views have publicly indicated an awareness of the fact that many standard aspects of contemporary cosmology, i.e. not only the more suspect elements such as the multiverse hypothesis or string theory, were explicitly condemned by Popper who, in 1994, described himself "a disgusted opponent" of the Big Bang theory (Kragh, this volume, Sect. 2.4).

In this chapter, I will examine whether Popper's scathing remarks about the methodology of cosmology could be moderated by the increasingly accepting attitude toward "metaphysical," i.e. non-testable, ideas in science, which appear especially in his later writings. Are there untestable ideas in cosmology that even a Popperian should be able to tolerate and what kind of problems are they meant to solve?

According to Popper (1982, 161), problem-situations in science are usually due to three factors:

Factor 1: "*the discovery of an inconsistency within the ruling theory*"

---

[1] In addition to the examples listed by Kragh, see Ellis and Silk (2014).

A. Lazutkina (✉)
LUT University, Lappeenranta, Finland
e-mail: a.k.lazutkina@gmail.com

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_6

Factor 2: “*the discovery of an inconsistency between theory and experiment—the experimental falsification of the theory*”
Factor 3: “*the relation between the theory and what may be called the ‘metaphysical research programme*’”

I will examine how the problem-situations are exemplified in contemporary cosmology. My discussion will mainly focus on instances of Factors 2 and 3. I will deal with these in reverse order, first considering, in Sect. 2, the notion of *metaphysical research programmes* (MRPs). In Sect. 3 I will describe the currently untestable ideas in contemporary cosmology and discuss whether at least some of them could be considered to collectively constitute a MRP. In particular, I will focus on the *Cosmological Principle* as a fundamental assumption of the Friedmann–Lemaître–Robertson–Walker family of models.

I will then consider the problem-situation related to the hypothetical *dark matter* from the perspective of both Factor 2 and Factor 3. This is because it is an auxiliary hypothesis,[2] designed to save the standard model of cosmology from being refuted, as well as an example of an untestable, “metaphysical” idea that could be seen to partially constitute a MRP for cosmology.

The conclusion I will draw from these considerations is that theories in cosmology, *when conceived of as the study of the whole universe*, remain on the untestable side of the demarcation criterion, and Popper is therefore consistent with his own views in not regarding such theories as scientific. Instead, cosmological models thus conceived, *fit the criteria of a MRP* as described by Popper, and could therefore have at least a heuristic importance for physics. However, cosmology *conceived of more modestly as the study of the largest-scale structures in the observable universe*, has produced testable and even well corroborated theories, which conform very well with Popperian methodology.

Finally, I will show that one does not have to be a Popperian in order to draw similar conclusions about the state of cosmology. I will use other methodological tools, namely Voishvillo’s (2003) reformulated, generalized *correspondence principle* and Niiniluoto’s (1987, 1999) *measures of truthlikeness*, to evaluate two different theories, the standard model of cosmology (ΛCDM[3]) and Modified Newtonian Dynamics (MOND),[4] as solutions to the mass discrepancy problem with regard to the internal velocities of the satellite galaxies of the Milky Way. These methodological tools, while not strictly Popperian, are either descendants of his ideas (truthlikeness) or are at least motivated by scientific realism (correspondence principle).

---

[2] Properly speaking, though, dark matter is not one hypothesis but at least five, as Merritt (2020, 14, 155) points out. I will here treat these hypotheses as one for simplicity.

[3] The full name of the model is the Lambda Cold Dark Matter model where the “Lambda” refers to the cosmological constant, or dark energy, and “Cold” to the type of non-baryonic dark matter particles postulated in the model.

[4] MOND is often characterized as a theory of modified gravity, but it is perhaps best described as a research programme at the heart of which rests Milgrom’s law, which can be interpreted either as a modification of the law of gravitation or the law of inertia. For an outline of MOND, see Sect. 4.1.

2 Metaphysical Research Programmes

Popper is famously known for holding that the demarcation between scientific theories and non-scientific theories, such as metaphysics and pseudo-science, is determined by their testability, but he did not put pseudo-science and metaphysics (at least not all of it) in the same basket. By 1934, in *Logik der Forschung*, he already held the view that "influential metaphysics" had heuristic importance in scientific theorizing (see Lakatos 1969, 178). Popper's views evolved during the subsequent decades, and in the "Metaphysical Epilogue" of Volume 3 of his *Postscript to The Logic of Scientific Discovery* (published in 1982), he describes science as almost always being "under the sway of metaphysical–that is, untestable–ideas; ideas which not only determine what problems of explanation we shall choose to attack, but also what kinds of answers we shall consider as fitting or satisfactory or acceptable, and as improvements of, or advances on, earlier answers." (ibid., 161) According to Popper, these ideas are organized into *metaphysical research programmes* (MRPs), which contain "general views of the structure of the world," "general views of the problem situation in physical cosmology," and "together with a view of what the most pressing solutions are, a general idea of what a satisfactory solution of these problems would look like." (ibid.).[5]

Popper goes on to list the ten MRPs that he considers to have been the most important in terms of their influence on physics:

1. the "Block Universe" of Parmenides
2. the "Atomism" of Leucippus and Democritus
3. the "Geometrization" of the Pythagoreans, Plato, Eudoxus, Callippus, and Euclid
4. the "Essentialism and Potentialism" of Aristotle
5. the "Renaissance" physics of Copernicus, Bruno, Kepler, Galileo, and Descartes
6. "The Clockwork Theory of the World" of Hobbes, Descartes, and Boyle
7. the "Dynamism" of Newton, Leibniz, Kant, and Boscovich
8. the "Fields of Forces" of Faraday and Maxwell
9. the "Unified Field Theory" of Riemann, Einstein, and Schrödinger
10. "The Statistical Interpretation of Quantum Theory" of Born (ibid., 162–164).

He makes the claim that, although the central ideas of these programmes were not testable (and some are even currently not testable), they were *criticizable*, as is evidenced by the fact that there was a progression of ideas criticized on theoretical grounds, and replaced by new ideas (ibid., 172). The last two programmes (9 and 10) contain ideas that contradict each other and give rise to what Popper calls a *schism in physics*: "Instead of a problem situation within a research programme, or relative to

---

[5] There are obvious comparisons to be made with Popper's notion of a MRP with Lakatos' *scientific research programmes*. Indeed, Lakatos was greatly indebted to the work of Popper, Agassi and Watkins in this regard (Lakatos 1969, 177–178). Although I cannot pursue that connection here, see Sect. 3.2 for a discussion of some parallel ideas.

a research programme" there is "a clash between two research programmes, neither of which seems to be doing its job." (ibid., 173) Specifically, the schism concerns the *interpretation* of classical physics and of quantum theory. MRP 9 describes all matter as disturbances or vibrations of geometrized fields, but MRP 10 takes an instrumentalist view of those fields, which now represent purely statistical probabilities of finding a particle in a certain state and position (ibid., 164). While classical physics was often interpreted in a determinist and objective way, in quantum theory this leads to highly counterintuitive consequences, which have led many to abandon the objective interpretation, or worse, lose interest in interpreting physical theories altogether.

It is safe to say that since the publication of *The Postscript*, this schism has not been resolved. If cosmology is simply understood as a branch of physics, this would seem to preclude the examination of cosmology's problem-situation in relation to a MRP. However, cosmology, at least when conceived of as the study of the entire universe, differs from other areas of physics in notable ways, due to the uniqueness of its object of study, as well as inherent difficulties in obtaining knowledge about regions of the universe to which we lack observational access. However the schism between MRP 9 and MRP 10 will be resolved, if it is resolved, the cosmological project, as it has been conceived by most of the research community, has required the adoption of several untestable assumptions that guide cosmological research. It is in this sense that cosmology has its own problem-situation in relation to a MRP.

2.1 Popper's Use of the Term "Metaphysical"

Before further examination of contemporary cosmology in light of the notion of a MRP, some remarks are in order about Popper's equation of "untestable" with "metaphysical," which contemporary cosmologists understandably might not welcome as a characterization of their research. Firstly, as Popper himself (1983, 74), notes, this is a technical term in his use.[6] Furthermore, as a scholar of the history of philosophy, Popper was obviously aware of different definitions of metaphysics, and was opposed to essentialist definitions at any rate, since he did not think science and philosophy should be in the business of answering "What is?" type of questions, but instead ought to focus on solving problems (see Ribeiro 2014, 209).

Nevertheless, many philosophers have objected to Popper's conception of metaphysics, recent examples being Akrami (2009) and Ribeiro (2014). Ribeiro argues, over Popper's conception, for what Popper himself calls the traditional way of defining metaphysics as "general theories about the nature of the world." By Popper's definition, theories as different as the germ theory of disease on the one hand, and Plato's Theory of Forms on the other, are all examples of metaphysical or formerly metaphysical theories. In contrast, Ribeiro proposes non-testability as a necessary but

---

[6] See also Lakatos (1969, 168, n. 58).

not sufficient criterion for a criticizable theory to be considered metaphysical. Metaphysical theories must also be sufficiently *general*. In fact, according to Ribeiro, the non-testability of metaphysical theories *follows* from their high level of generality.[7] Instead, non-general and untestable theories ought to be considered *speculative or proto-science* if they are criticizable, or *pseudo-science* if they are not criticizable.

Ribeiro claims, perhaps plausibly, that Popper was more interested in demarcating science from pseudo-science, and not in demarcating either science or pseudo-science from metaphysics, and thus overlooked the criterion of generality. She points to Popper's own wavering between describing MRPs in physics as being constituted by metaphysical ideas on the one hand, and "speculative physics" on the other (Popper 1982, 161–162), as "telling" of the fact that there is a conflation of two types of ideas in Popper's use of the term "metaphysical."

Ribeiro concludes that equating "metaphysical" with "untestable" is simply too confusing. For her, there is no reason, apart from being able to maintain the Popperian demarcation criterion, why we should not prefer the traditional definition of metaphysics, if as Popper himself claims, the boundary between science and metaphysics is blurry anyway.

The terminological disagreement between Popper and Ribeiro need not be resolved here. Firstly, both seem to think there are almost always untestable ideas, of varying levels of generality, in the background of scientific theories, which would make the boundaries between science and proto-science, as well as between science and metaphysics, blurry. Secondly, both seem to think there is progress from the untestable ideas, both general (e.g. atomism) and non-general (e.g. germ theory of disease), to scientific theories. Thirdly, it is not clear to me that the untestable ideas present in contemporary cosmological theories that I am concerned with here are of a sufficiently high level of generality to be metaphysical in the traditional sense. Certainly none of them are sufficiently general, such that their non-testability follows simply from their generality. I only wish to highlight the possibility of choosing different terminology for those to whom the term "metaphysical" in "metaphysical research programmes" would be upsetting in the context of contemporary cosmology, or for those who object to Popper's conception of metaphysics for other reasons. Although in what follows, I will stick to Popper's terminology, in my view one could equally well switch terminology and call MRPs speculative or proto-scientific research programmes in the context of contemporary cosmology.

3 Metaphysical Ideas in Contemporary Cosmology

We are now in a position to examine contemporary cosmology through the prism of Popper's notion of MRPs. Are there untestable principles or ideas in cosmology

---

[7] Although I cannot discuss the point here, it is an interesting question whether this claim is defensible. Certainly, it is not obvious that a highly general theory, such as materialism, *must be* untestable.

that “determine what problems of explanation we shall choose to attack, but also what kinds of answers we shall consider as fitting or satisfactory or acceptable, and as improvements of, or advances on, earlier answers”? To ask this is to probe at the foundations of cosmology.

One has to begin from the fact that, at astronomical scales, gravity is the dominant interaction, so a theory of gravity is the starting point of a cosmological model. The field equations of Einstein’s general theory of relativity are considered the default choice in this regard.[8] In a classical case of underdetermination, these equations allow for a wide range of possible cosmological models, so assumptions must be added and observational evidence taken into account to restrict the range of possibilities. Perhaps the most important of these assumptions is known as the Cosmological Principle (CP), which, following Jung and Beisbart (2006), I shall define here as the claim that, at any time, the universe is homogeneous[9] at sufficiently large scales.[10]

While Popper was dismissive of the CP (Kragh, this volume, Sect. 2.2), I will propose here that the CP is in fact the main component of what in a Popperian sense constitutes a MRP for cosmology. To clarify this point, I must now provide two contrasting outlines of cosmology as a field of study. Ellis (2006, 1183) defines cosmology as “the study of the large-scale structure of the Universe, where ‘the Universe’ means all that exists in a physical sense,” whereas observational cosmology “aims to determine the large-scale geometry of the observable universe and the distribution of matter in it from observations of radiation emitted by distant objects.” Ellis sees observational cosmology as a subdiscipline of cosmology, whereas Beisbart (2009) highlights the possibility of looking at these as two alternative conceptions of the discipline. Cosmology as the study of the universe as a whole is an ambitious project, whereas studying the large-scale structures of the observable universe is a more modest one.

Assuming the CP in the respective contexts of these two projects, i.e. for the observable universe and for the entire universe, are two very different things. With regard to the observable universe it is in principle observationally refutable *and* verifiable (since it is not a universal principle), and there is at least considerable evidence in its favor (see for example Lahav 2001; Beisbart 2009; Sarkar et al. 2009;

---

[8] In Sects. 3.2, 4.1, 4.2 and 4.3 I will look at MOND as an alternative to assuming the universal correctness of general relativity, but only in a limited sense, as it applies to the dark matter hypothesis. The foundations of a Milgromian cosmology would be deserving of a much thorough treatment than I could provide here.

[9] Roughly, homogeneity is uniformity with respect to location. The CP is often thought to include the claim that the universe is isotropic (isotropy is, roughly, uniformity with respect to direction), but the isotropy of the observable universe is testable (and therefore no principle has to be assumed for the claim), and global isotropy follows analytically from global homogeneity and the isotropy of the observable universe (Jung and Beisbart 2006, 252).

[10] See Butterfield (2014, 61) for different approaches to defining “sufficiently large scales.” It is common to confuse the CP for another principle, known as the Copernican Principle, according to which our position in the universe is not “privileged” or “special.” Jung and Beisbart (ibid.) remark that while isotropy and homogeneity are mathematically defined concepts, “privileged” and “special” have no such unambiguous meaning. They have also shown the logical gap between two principles: the Cosmological Principle implies the Copernican Principle, but not vice versa.

Yang and Saslaw 2011; Maartens 2016). For the entire universe, Popper's skeptical attitude toward the testability of the CP finds representation among contemporary philosophers of cosmology, such as Beisbart (2009) and Butterfield (2014), but this does not contradict my suggestion that it partially constitutes a MRP. To elaborate on this suggestion, I must now look at the motivation for the adoption of the CP and its theoretical and interpretive roles in cosmology.

3.1 The Cosmological Principle as a Constituent of a MRP

Since my primary aim here is not to offer a historical account, I will mention only some key developments.[11] In 1917, Einstein adopted the idea of a homogeneous universe for his first cosmological model to satisfy Mach's principle (Torretti 2000, 171), as well as for mathematical convenience.[12] Since then, several other considerations have entered the picture. In the 1920s Alexander Friedmann demonstrated that one can use the CP to build a coordinate system in order to solve Einstein's equations for a dynamical model of an expanding universe. (Ntelis 2017, 2) With evidence of the expansion of the universe taken into account, a class of models known as the Friedmann–Lemaître–Robertson–Walker (FLRW) models became the mathematical basis for realistically describing the observable universe. Butterfield (2014, 61) observes how radical the notion of allowing the geometry and material contents of the universe to change over time was initially, but is now considered one of the main motivations for accepting the CP. Butterfield (ibid.) considers the CP a "lucky break" for avoiding underdetermination in cosmology due to its mathematically elegant consequences for the spacetime metric, its mathematical relation to other principles, and the aforementioned fact that there is considerable evidence that it holds with regard to the observable universe.

Returning to the two alternative conceptions of cosmology, we may now say that Popper's harsh criticism of cosmology does not seem to apply to the more modest project of describing the largest observable structures of the universe, at least as far as the claims about homogeneity and isotropy go.[13] The difficulties with the CP begin when claims are made about the universe as a whole, since there is no straightforward observational evidence we could appeal to. Here astronomers, cosmologists and philosophers have traditionally relied on some type of "fair sample" hypothesis, according to which the universe as a whole exhibits the same properties as the regions we are able observe. But why should we think this is the case?

---

[11] See Kragh (1996) for a historical account.

[12] Convenience undoubtedly still motivates the adoption of the CP, but as Jung and Beisbart (2006, 251) ask, "Why should Nature facilitate our calculations?"

[13] This does not preclude the possibility of criticizing other aspects of cosmology, such as its reliance on the untestable auxiliary hypotheses of inflation, dark matter and dark energy. This is discussed further in Sect. 3.2.

Beisbart (2009: 189–201) examines several strategies for justifying this assumption:

One strategy is to argue that it is more likely than not that initial conditions compatible with the observable universe would lead to a universe that obeys the CP globally, but this has not been established, and would be difficult to establish due to there being "no natural measure for initial conditions from which probabilities can be obtained." (Ibid., 193).

A second type of strategy is to argue that models that conform to the CP have greater explanatory power than those that do not. In particular, inflationary cosmology is thought to provide such a model, but this line of thought meets the following difficulties:

1. there are inflationary models that result in a universe in which the observable universe obeys the CP but other regions do not;
2. the purely hypothetical object known as the inflation field is an ad hoc maneuver to prevent the refutation of the standard model, and therefore methodologically suspect (see also Merritt 2020, 39);
3. too much hangs on the "style" of explanation preferred (Beisbart 2009, 196).

A third strategy is to attempt to generalize from the assumed invariance of physical laws within the universe, to the invariance between physical magnitudes within the universe. It suffices to say that this is a logical leap that would require further argumentation.

A fourth strategy rests on an induction made from the observable universe to regions beyond it, but it is not clear what kind of inductive approach could work here. For example, in a Bayesian approach, there is no way to fix the prior probability of the universe being homogeneous (ibid., 200). While Beisbart concludes that no attempt is successful at the moment, he emphasizes that this may change depending on future observations.

From a Popperian standpoint we might ask: why not merely assume as a working hypothesis that the CP holds globally, and attempt to formulate testable consequences of this hypothesis? Jung and Beisbart (2006, 246–247) suggest that the best we can hope for is to check for consistency with other well established theories. However, there are cosmological models that violate the global CP but describe the observable universe realistically (for a review, see Sundell 2016), so this only brings us back to the problem of underdetermination.

I must come to the conclusion that there is no compelling evidence for assuming the CP for the entire universe, and assuming it for the entire universe does not result in unique predictions for the observable universe. However, it has guided cosmology for the past 90 years (Beisbart 2009, 176), provides constraints for initial conditions (ibid., 201) and affects the way light propagation is studied (Jung and Beisbart 2006, 246) (just to mention a few consequences for modeling). An independent result that *would* confirm or refute the CP would, thus, be a significant step forward in cosmology. Taking the assumption of a homogeneous universe to be a constituent of a MRP for cosmology, this is precisely what one would expect:

> By raising the problems of explanation which the theory is designed to solve, the metaphysical research programme makes it possible to judge the success of the theory as an explanation. (Popper 1982, 161)

3.2 The Standard Model of Cosmology (ΛCDM) as a Metaphysical Research Programme

I am now in a position to suggest that there is a MRP in cosmology and that the CP is a part of it. But what exactly is that MRP? I have only looked at the CP so far, but other claims about the universe as a whole generally face the same challenges as the CP does, and can hence be characterized as metaphysical in the Popperian sense. Therefore, I tentatively propose that any sufficiently developed and stable[14] cosmological model, when *cosmology is conceived of as the study of the whole universe,* could be considered a MRP. This formulation is vague (what counts as "sufficiently developed and stable"?), but then, Popper does not provide any strict criteria for a MRP, and essentialist definitions are not Popperian in spirit, anyway.

I also say "could be considered," since whether we *should* examine anything through the notion of a MRP depends on whether this is fruitful for understanding the phenomenon in question. I certainly think it *is* useful in the case of some of the untestable features of the standard model of cosmology, ΛCDM, since they are instrumental in defining the problem situation that any theory has in relation to it.

Exactly which features one would include in the MRP depends on how strict the requirement for stability is. Although it is not my primary purpose here to compare Popper's notion of MRPs to Lakatos' (1969) notion of scientific research programmes (SRPs), it is worth noting a parallel: Lakatos states that the "hard core" of a SRP can develop slowly in some case, "by a long, preliminary process of trial and error" (Lakatos 1970, 48, note 4; as cited by Merritt 2020, 30). Merritt (ibid., note 10) mentions the hypothetical dark matter in this parallel context: it has been a feature of the standard model for about 40 years and most cosmologists present its existence as a known fact (despite no independent evidence of its existence), so it could be reasonably included in the MRP.

The fact that MOND, the main rival of the dark matter hypothesis, is not considered by standard model cosmologists to be a promising answer to the so-called problem of missing mass despite its numerous successes also speaks in favor of including dark matter in the MRP, since this points to clear criteria for "what kinds of answers we shall consider as fitting or satisfactory or acceptable, and as improvements of, or advances on, earlier answers." (Popper 1982, 161).[15]

---

[14] By this I mean stable over time in terms of the ideas it contains. Popper's examples of MRPs contain ideas that in some cases were held for centuries, in some cases less.

[15] This is not an endorsement of the current situation. While some aspects of Popper's methodology of science, such as the demarcation criterion, are prescriptive, I take his claims about the significance of the MRP to be largely descriptive, and this is how my claims about the MRP in cosmology should also be read.

There are several important differences between the CP and the dark matter hypothesis as parts of the MRP. One of these concerns the scope of their falsifiability. The CP is unfalsifiable for the entire universe, whereas the dark matter hypothesis is unfalsifiable simpliciter. It is perhaps tempting to assume the CP for the whole universe, since it has been confirmed to correctly describe the observable regions. But what makes dark matter so appealing? The typical answer would be that there is no serious alternative, reflecting its aforementioned role in defining the problem situation. Contra this, Merritt (2020) has provided serious considerations in favor of an alternative, known as MOND, from a Popperian-Lakatosian perspective.

Additionally, Merritt (2017) has shown that standard model cosmology has features of what Popper calls conventionalism, i.e. ad hoc stratagems are used to avoid the refutation of the standard model. Over-reliance on these is an indication of a degenerating programme, although Merritt refrains from stating whether the programme has degenerated beyond hope.

Is the use of ad hoc stratagems problematic, if elements of the standard model are viewed as a MRP? After all, there is no requirement of refutability in metaphysics. I wish to re-state here that Popper's conception of metaphysics does not coincide with his conception of pseudo-science. Although the metaphysical ideas included in the MRP are untestable, Popper sees them as "speculative physics, or perhaps as speculative anticipations of testable physical theories" (Popper 1982, 161). They must therefore be criticizable unlike pseudo-scientific theories, which according to their proponents, are constantly verified no matter what.[16] Whether auxiliary hypotheses such as dark matter are genuinely criticizable depends not only on the nature of the hypothesis but also on the attitudes of the research community. As a worrying example, Merritt (2017, 47) reports how no graduate level cosmology textbooks even mentions the empirical mass discrepancy–acceleration relation (which is commonly thought to hint at the breakdown of Newtonian gravity at low accelerations rather than the presence of undetectable dark matter).

As Kragh (this volume) has documented, not all standard model cosmologists have received the criticism of their colleagues with gratitude, let alone when it is seen to originate from the prescriptions of philosophers such as Popper. As an additional example, de Swart, Bertone and van Dongen (2017, 6) complain that Popperian critiques of standard model cosmology do not capture the rational motivation for accepting the dark matter hypothesis as practically confirmed. Instead their roughly sketched argument amounts to suggesting that we need to better understand the "actual practice and methods of physics, astronomy and cosmology" (ibid.). But this is hardly a good response to someone who is criticizing the *actual practice and methods* of standard model cosmologists, especially when these critics include astrophysicists and cosmologists who understand these practices and methods very

---

[16] As an example of what the fruitful interplay of metaphysics and physics can look like, Popper (1982, 165–173) describes how, during the seventeenth, eighteenth and nineteenth centuries, there was genuine progress from MRP 6 ("The Clockwork Theory of the World") to MRP 7 ("Dynamism") and then to MRP 8 ("Fields of Forces") largely on theoretical grounds.

well.[17] One day, if their programme has already yielded genuine discoveries,[18] standard model cosmologists might be in the right to complain about methodological prescriptions—not before.

4 Beyond Popper

In the previous sections, my focus has been on applying Popper's methodology to contemporary cosmology. The considerations in the following sections are not radical departures from Popper's ideas, but are motivated by ideas that are either directly descended from Popper's thought, or at least motivated by a similar critical realist approach to methodology. The motivation behind this is to show that one does not have to be a Popperian to be critical of ΛCDM and to take MOND to be a serious rival to it. In order to illustrate this, I will use the methodological tools of *principle of correspondence* and *measures of truthlikeness* to examine these two rivals. 4.1 and 4.2 examine MOND only, whereas 4.3 compares MOND and ΛCDM directly.

4.1 A Brief Outline of MOND

> In science, problem situations are the result, as a rule, of three factors. One is the discovery of an inconsistency within the ruling theory. **A second is the discovery of an inconsistency between theory and experiment - the experimental falsification of the theory.** The third, and perhaps the most important one, is the relation between the theory and what may be called the 'metaphysical research programme'.
>
> (Popper 1982, 161; emphasis added)

The Newtonian predictions for the rotational velocities of objects at the edges of galaxies do not match our observations. Two options for correcting this discrepancy are the introduction of a hypothetical object or modifying the Newtonian laws.[19] I have already discussed the dark matter hypothesis as an example of a hypothetical object and will now briefly discuss MOND as an alternative solution.

---

[17] In addition to Merritt's work, see Kroupa (2012) for an astrophysicist's analysis of the repeated falsifications of the standard model.

[18] In the present case, an example of a genuine discovery would be the independent detection of a (class of) dark matter particle(s), and a successful study of its/their properties that would explain the observed Milgromian dynamics at the edges of galaxies.

[19] See Lazutkina (2017) to see how MOND and dark matter can be compared to other cases in the history of astronomy and physics.

Developed by Mordehai Milgrom in 1983, MOND describes the difference in the dynamics of objects that depends on whether the objects are situated in high-acceleration regimes, such as objects orbiting the Sun in our Solar System, or low-acceleration regimes, such as objects at the edge of our galactic disk orbiting the center of our galaxy (Milgrom 1983a, b, c).

High-acceleration regimes are also known as Newtonian regimes, because they were the only ones observed in detail before the work of Zwicky, Rubin, and others that lead to the observations that contradict the prediction, which follows from the conjunction of Newton's second law and Newton's law of gravitation.[20]

By noticing that the contradiction follows from the conjunction of these two laws, Milgrom (2014) suggests that a modification of either is possible. Therefore, the core of MOND, known as Milgrom's law, is not strictly speaking a theory of modified gravity nor modified inertia, but rather accounts for the empirical dependence between acceleration and dynamical behavior in a way that can be interpreted as a modification of either. Milgrom's law thus implies the disjunction of modified gravity and modified inertia, although it is silent on how to construct a full theory of either type.

Although MOND requires a relativistic extension as the basis of a realistic cosmological model, it is a research programme that has steadily produced unique, novel predictions that have been corroborated or confirmed, as will be exemplified in Sect. 4.3 (see also Merritt 2020, 194 for a summary of MOND's successes). It also passes the methodological test of conforming to the principle of correspondence, as I will next demonstrate in Sect. 4.2. MOND has its problems of course, but they are not necessarily insurmountable. Recently, the development of a relativistic extension of MOND known as RelMOND was able to solve a long-standing problem for MOND, namely to reproduce the observed Cosmic Microwave Background (CMB) and matter power spectra (Skordis and Złosnik 2020). Achieving empirical adequacy in this regard is especially important for MOND because, so far, only the standard model has been able to do so, and this has been considered a significant advantage of the standard model over MOND.

4.2 The Correspondence Principle and MOND

When the prediction of a theory turns out to be false, and the ad hoc conventionalist stratagems mentioned previously are avoided, the theory is thereby refuted. Whatever new theory is proposed to take the place of the old theory must either agree with the empirically successful parts of the old theory, or if the old theory is completely discarded, there must be an explanation—founded on the new theory—for why the old theory had the limited empirical success it did.

---

[20] Strictly speaking one should already speak of modified versions of Newton's laws, because the scope of their application is already constrained by the conditions given by the general theory of relativity, whereas the unmodified, falsified Newtonian laws have no such restrictions.

This idea has its origin in a 1913 paper by Niels Bohr, although the term "correspondence principle" (*Korrespondenzprizip*) did not appear in his writings until 1920 (van der Waerden 1967, 7–8).[21] In a more general form, a philosophical formulation of it was given, among others, by I. V. Kuznetsov (1948, 56, translation mine): "Theories whose validity is experimentally established for a particular field of physical phenomena, are not eliminated as something false with the emergence of new more general theories, but retain their significance for the former field of phenomena, as the limiting form and special case of the new theories."

However, as shown by E. K. Voishvillo (2003), there are inaccuracies in this formulation. The old theory is not a special case of the new one, since it turns out to be false (in light of the refuting observation). Instead, a *modified* version of the old theory is a special case of the new theory. The statements of the old theory are reformulated by adding new conditions (in light of the new one), thereby narrowing the scope of its application, and deleting the implied false part from it. In the relevant fields of theoretical knowledge, the implementation of this procedure is a formal way of testing whether a new proposed theory fits the current scientific picture. Its failure to do so is a formal reason to discard it (ibid.).

I also follow Aliabadi (1996, 9–10, 45–55), who holds that the old theory should merely be a good approximation of the new one in limited cases.

Here, Voishvillo's approach will be applied to MOND in order to demonstrate that the modification of Newton's second law is a special case of the law of the general theory of relativity and at the same time a special case of one of the interpretations of the modification of this law by Milgrom $F = m\mu(a/a_0)a$.

Milgrom introduces a new constant, critical acceleration—a_0—($a_0 \approx 1{,}2 \times 10^{-10}$ m/c^2). When the acceleration of an object significantly exceeds this threshold, it obeys Newtonian dynamics, and when it is much lower than it, its behavior is accurately described by MOND. The transition from the Newtonian regime to the deep-MOND regime is described by an interpolating function, μ, which is currently unspecified, yet thought to be quite steep (Famaey and Zhao 2006). Thus, the dependence $\mu(a/a_0)$ is introduced. For large accelerations, the value of this term is 1, i.e. Newton's laws (F = ma) are preserved. For small accelerations, where a is less than a_0, we obtain $GM/r^2 = \mu(a/a_0)a$. Thus, Newtonian dynamics, with the addition of the condition $\mu\ (a/a_0) \approx 1$), becomes a special case of MOND.

According to Newton's second law:

$$\mathbf{F} = \mathbf{d}\,(\mathbf{mv})/\mathbf{dt},$$

i.e.

$$\forall x \forall v \forall m \forall t \forall f\,((\mathrm{V}(v, x, t)\ \&\ \mathrm{M}(m, x)\ \&\ \mathrm{F}(f, x, t)) \rightarrow f = \mathrm{d}\,(mv)/\mathrm{d}t)$$

---

[21] While philosophically opposed to Bohr's other famous principle, i.e. the principle of complementarity, Popper (1963, 101) considered the correspondence principle "extremely fruitful" for scientific research.

Where x is a body, and f, m, v, and t are real numbers—the possible values of force, mass, velocity, and time. V (v, x, t) means—the number v is the value of the speed of the body x at the time t, M(m,x) means—m is mass of x, F(f,x,t) means—f is force that acts on body x at moment t.

In MOND: $\mathbf{F = m\boldsymbol{\mu}(\boldsymbol{a/a_0})\boldsymbol{a}}$,[22]

The logical form of the law can be given thus:

$$\forall x \forall v \forall m \forall t \forall f((\mathrm{V}(v, x, t) \,\&\, \mathrm{M}(m, x) \,\&\, \mathrm{F}(f, x, t)) \rightarrow f = m\mu(a/a_0)a)$$

Let us introduce a special condition D, thereby narrowing the scope of application of Milgrom's law, and obtaining a special case of it. In order to consider Newtonian dynamics as a particular case of Milgrom's law, D: μ $(a/a_0) \approx 1$. From Milgrom's law an expression logically follows with the condition D introduced into the antecedent. Thus, this expression is a special case of Milgrom's law:

$$\forall x \forall v \forall m \forall t \forall f((\mathrm{V}(v, x, t) \,\&\, \mathrm{M}(m, x) \,\&\, \mathrm{F}(f, x, t) \,\&\, \mathrm{D}) \rightarrow f = m\mu(a/a_0)a)$$

This is equivalent to:

$$\forall x \forall v \forall m \forall t \forall f((\mathrm{V}(v, x, t) \,\&\, \mathrm{M}(m, x) \,\&\, \mathrm{F}(f, x, t) \,\&\, \mathrm{D}) \rightarrow (f = m\mu(a/a_0)a \,\&\, \mathrm{D}))$$

Since the condition D means μ $(a/a_0) \approx 1$, we obtain $f = \mathrm{d}\,(mv)/\mathrm{d}t$ as the consequent.

Thus, we have:

$$\forall x \forall v \forall m \forall t \forall f((\mathrm{V}(v, x, t) \,\&\, \mathrm{M}(m, x) \,\&\, \mathrm{F}(f, x, t) \,\&\, \mathrm{D}) \rightarrow f = \mathrm{d}\,(mv)/\mathrm{d}t)$$

This demonstrates that Newton's second law is a special case of MOND when condition D is taken into account, that is, when working with standard accelerations. This result proves that MOND satisfies the formal requirement posed to a theory that aims to succeed an old theory that is empirically successful within constraints. Namely, the modified version of Newtonian dynamics is true within the constraints given by condition D. Thus, a modified version of Newtonian dynamics becomes a special case of MOND.

4.3 Truthlikeness: ΛCDM Versus MOND

The notion of *verisimilitude* was introduced to contemporary philosophy of science by Popper. A part of his falsificationism and critique of inductivism, is the claim that we are never justified in claiming that a theory is true or even probably true. Yet, Popper was a scientific realist and, accepting the Tarskian notion of truth, claimed

---

[22] For simplicity, here and later, a is used instead of dv/dt for Milgrom's law.

that scientific progress can be understood as more *truthlike* theories replacing less truthlike theories.

According to Popper's definition of truthlikeness, known as the *content approach*, theory A is more truthlike than theory B if A has more truth content than B without implying more falsity content than B, where the content of a theory is understood as the set of claims it makes. Popper's approach works when A is true (has no falsity content), but David Miller (1974a, 1974b) and Pavel Tichý (1974) both independently proved that when a theory has some falsity content, its truth content cannot be increased without increasing its falsity content. A consequence of this is that, following Popper's approach, we cannot say that A is more truthlike than B, when A has some falsity content, and therefore cannot make sense of scientific progress (Oddie 2016).

Despite the problems with Popper's approach, the notion of truthlikeness has become an important part of scientific realism. Most agree that a good way to make sense of scientific progress is to say that, for example, general relativity is more truthlike than Newtonian dynamics. The concept of truthlikeness can also be used as part of a realist reply to the pessimistic meta-induction (in both its semantical and epistemological forms) and the problem of meaning variance.

Despite the intuitive appeal of truthlikeness, the question of specifying the notion in a coherent way remains. There are various competing approaches, and it would be impossible to provide a comprehensive survey of them here. Instead, one particularly promising approach will be selected for closer examination, namely the *likeness approach*. Niiniluoto (1999, 68) summarizes this approach in the following way: truthlikeness = truth + similarity. According to this approach, the measuring of truthlikeness is relative to what he calls *cognitive problems*, which are represented by either finite or infinite sets of statements, expressed in an interpreted and semantically determinate language, whose elements are mutually exclusive and jointly exhaustive possible answers to the problem. A single element represents a *complete* potential answer to the cognitive problem, whereas a disjunction of several elements represents a *partial* potential answer (ibid).

Although the measure of truthlikeness must be specified for each specific, concrete cognitive problem, Niiniluoto provides measures for some "canonical" cognitive problems. The simplest type of cognitive problem is a yes/no question. From the point of view of applying truthlikeness to astrophysics and cosmology, the more interesting types of cognitive problems concern the magnitude of some physical quantity (e.g. the mass of a star), or the functional relation of some physical quantities (e.g. the dependence between the distance of a star from the galactic center and its rotational velocity) (ibid., 69). The latter kind of measure is of special interest to us, since the theories discussed in the previous chapters are motivated by the discrepancy between functional dependencies derived from empirical data and theoretical predictions. Theories in astrophysics imply statements regarding functional dependencies between observable physical quantities. Typically, these statements are only approximately accurate at best, and so strictly speaking each of them is false, assuming that our measurements of the quantities correspond to their true values. This is why the measure of truthlikeness, which is suitable for cognitive problems relevant to astrophysics and cosmology, cannot be expressed by a measure of the true and false sentences implied by physical theories (ibid., 73).

Rather, the measure of truthlikeness provided by Niiniluoto to cognitive problems concerning point values and functional dependencies is founded on abstractions of the properties of the Euclidean plane, known as *the metric space*. For a measure of truthlikeness to count as a metric in this sense, it must satisfy strict formal conditions. However, so-called *distance functions* are able to preserve the relevant features of metrics, if we are not interested in the numerical value of the metric but the results it gives for comparative purposes (Niiniluoto 1987, 1–4).

It is precisely these comparative results that are valuable in the present context. With the assumption that our measurements of the relevant physical quantities are accurate, it is possible to compare the truthlikeness of different theories implying functional dependencies between the quantities, by employing a metric known as the *Minkowski distance* (Niiniluoto 1999, 69).

To calculate the distance between the two points values, Niiniluoto (ibid.) provides the following equation:

$$\mathrm{d}\,(\mathrm{x},\ \mathrm{y}) = |\,\mathrm{x} - \mathrm{y}|$$

To calculate the distance between two functions, the equation is as follows:

$$\mathrm{d}\,(\mathrm{f},\ \mathrm{g}) = \int |\,\mathrm{f}\,(\mathrm{x}) - \mathrm{g}\,(\mathrm{x})\,|\,\mathrm{dx}$$

Since the measuring of truthlikeness is relative to a concrete cognitive problem, we will here consider the internal velocities of specific dwarf galaxies. One reason to choose this concrete cognitive problem is that postulating the dark matter hypothesis was originally motivated by the shape of galactic rotation curves, i.e. the velocities are much higher than what is predicted by Newtonian dynamics. This discrepancy can be formulated in terms of truthlikeness.

While Popper and Niiniluoto are interested in explaining the growth of scientific knowledge, I will here repurpose the formal apparatus of Niiniluoto's approach and use it as a heuristic methodological tool. The empirical data is assumed to be accurate (the truth), and the ways to get closer to the truth is to either introduce a hypothetical object or modify the theoretical predictions i.e. the theory of gravity. The goal, no matter which option is chosen (e.g. ΛCDM or MOND) is to try to minimize the distance between the predicted values and observational data, and thus to get closer to the truth.

The internal velocities of dwarf galaxies provide an excellent test for MOND and ΛCDM, because they have a low surface brightness (indicating low stellar mass) and high rotational speeds (indicating a high dynamical mass or the breakdown of Newtonian dynamics) (Strigari et al. 2008) Relevant data concerning their internal dynamics is available for nine dwarf spheroidal galaxies, which orbit the Milky Way galaxy. These are: Draco, Sculptor, Sextans, Fornax, Leo I, Leo II, Canes Venatici I, Carina, and Ursa Minor.

Based on the distribution of the visible matter of these galaxies, MOND predicts the rotational velocity of these galaxies. With ΛCDM, the story is more complicated: since it involves free parameters, ΛCDM makes no unique prediction regarding their internal dynamics. Instead, the hope of physicists working in this paradigm is to one day provide a theory of galaxy formation involving the gravitational interaction of baryonic matter and non-baryonic dark matter, which will explain the observed dynamics. The best that ΛCDM can provide at the moment is a post hoc simulation of the dynamics of these galaxies, and it is the truthlikeness of this that we can measure relative to our observations.

The results show that in 6 of 8 cases, MOND produces predictions (Alexander et al. 2017) closer to the truth than ΛCDM post hoc simulations (Fattahi et al. 2016), without requiring nearly as many free parameters:

Galaxy	Observed velocities	Predictions MOND	Post hoc sim. ΛCDM	Truthlikeness MOND	Truthlikeness ΛCDM
Fornax	20.1	20.8	25.5	0.59	0.16
Carina	11.3	9.9	13.8	0.42	0.29
Leo I	15.8	15.9	16.2	0.9	0.71
Leo II	11.3	11.6	12.8	0.77	0.4
Sculptor	15.8	14.9	15.7	0.53	0.9
Draco	15.6	15.1	14.7	0.67	0.53
Sextans	13.5	11.8	18.2	0.37	0.18
Ursa Minor	16.3	15.4	16.6	0.53	0.77

As the result of the measurement of truthlikeness is relative to concrete individual internal velocities of dwarf galaxies, we cannot say anything about the general success of these theories in terms of truthlikeness with regard to their ability to conform to observational data of these circular velocities in general. It is only possible to offer truthlikeness of the concrete predictions for concrete galaxies.

Nevertheless, this is but the first step taken in the application of the notion of truthlikeness to astrophysical theories. Further work must be done in order to compare the truthlikeness of these theories more generally.[23]

5 Conclusions

In this chapter I have analyzed some problem-situations in cosmology through the prism of Popper's notion of MRPs. I have examined a foundational principle of cosmology, the Cosmological Principle (CP), which states that the universe is homogeneous at sufficiently large scales. The result of the analysis is that the CP can be

---

[23] To see how the calculations are fully worked out, as well as truthlikeness measures for MOND predictions regarding the rotation curves of other galaxies, see Lazutkina (unpublished).

considered one of the main constituents of a MRP in contemporary cosmology. Dark matter also seems like a plausible candidate for inclusion in the MRP. The notion of a MRP turns out to be a fruitful point-of-view, because it clarifies the theoretical structure of contemporary cosmology: overall, cosmological theories, when the discipline is understood as the study of the whole universe, seem more like MRPs than scientific theories, if Popperian standards are applied, whereas cosmological theories, when the discipline is conceived of as the study of the largest-scale structures of the observable universe, can be scientific in principle. But even metaphysical ideas must be criticizable in Popper's opinion. The dark matter hypothesis, for example, informs the problem-situation to such a degree that some empirical evidence hinting in other directions is practically discarded, such as the mass-discrepancy-acceleration relation. The criticizability of the MRP that informs the standard model does not depend only on how the hypotheses are formulated at any one time, but on what the response to rational criticism is.

The response to criticism of the currently favored model depends, of course, on whether there exist viable alternative theories. In the present case, Modified Newtonian Dynamics (MOND) has been established as a successful research programme with a progressive problem-shift by its proponents within the disciplines of astronomy and cosmology (see esp. Merritt 2020 for a full-length treatment of the issue).

I have applied additional methodological tools, namely the *correspondence principle* and *measures of truthlikeness*, with the results favoring the viability and empirical adequacy of MOND. These considerations go beyond Popper's methodological prescriptions, but are descended from Popperian ideas, or at least motivated by a kind of critical scientific realism. MOND has been shown to be not only a viable alternative by conforming to the correspondence principle, but superior to ΛCDM in some respects. Not only does MOND adhere to Popper's methodological prescriptions unlike ΛCDM, but it is also more truthlike with regard to the concrete cognitive problems presented here.

The dark matter controversy is only one piece of the puzzle. The more fundamental issue is that, like ancient Greek atomists, contemporary cosmologists are far away from being able to test their theses about the universe as a whole. One might even say that we are currently much further away from the testability of these claims than the ancient Greeks were in relation to modern atomic theory. Then again, how conceivable would modern scientific instruments and experimental techniques have been to the atomists 2500 years ago? MacIntyre (1981/2007, 93) attributes to Popper the idea that radical future innovations are impossible to predict, because the prediction involves the conception of the innovation itself. Hence, we are not in a position to conclusively predict whether the CP, for instance, might one day become testable.

This is not to say that Popper's harsh words against contemporary Big Bang cosmology are not understandable in light of his methodological views. Many features of the standard model are presented as proven fact.[24] Some cosmologists and philosophers of cosmology seem to acknowledge the methodological limitations of cosmology, and the tentativeness of the current favored model on the one hand,

---

[24] For a notable and egregious example concerning the status of dark matter, see Clowe et al. (2006).

but also speak of the "successes" and "discoveries" of the discipline, as if these were settled matters.[25]

Instead of hanging on to a degenerating programme, a cosmologist adhering to Popperian norms would take a step back, acknowledge the problems of Big Bang cosmology for what they are instead of hailing them as "discoveries," and take an attitude of epistemic humility together with the freedom of making bold conjectures from which he might one day hope to derive testable consequences.[26]

References

Akrami, M.: The difficulties with Popper's nontraditional conception of metaphysics. In: Parusniková, Z., Cohen, R.S. (eds.) Rethinking Popper. Boston Studies in The Philosophy of Science, vol 272. Springer, Dordrecht (2009) https://doi.org/10.1007/978-1-4020-9338-8_30

Alexander, S., et al.: MOND calculations of bulk dispersions and radial dispersion profiles of milky way and andromeda dwarf spheroidal galaxies. Astrophys. J. **835**, 233–246 (2017)

Aliabadi, Y.: The correspondence principle and theory choice in physics. London School of Economics, London (1996)

Beisbart, C.: Can we justifiably assume the cosmological principle in order to break model underdetermination in cosmology? J. Gen. Philos. Sci. **40**, 175–205 (2009)

Butterfield, J.: On under-determination in cosmology. Stud. Hist. Philos. Mod. Phys. **46**, 57–69 (2014)

Clowe, D., Bradac, M., Gonzalez, A.H., Markevitch, M., Randall, S.W., Jones, C., Zaritsky, D.: A direct empirical proof of the existence of dark matter. Astrophys. J. Lett. **648**, L109 (2006)

de Swart, J., Bertone, G., van Dongen, J.: How dark matter came to matter. Nat. Astron. **1**, 0059 (2017). https://doi.org/10.1038/s41550-017-0059

Ellis, G.: Issues in the philosophy of cosmology. In: Butterfield et al. (eds.) Handbook of the Philosophy of Science: Philosophy of Physics, vol B. Elsevier, Amsterdam (2006)

Ellis, G., Silk, J.: Scientific method: defend the integrity of physics. Nature **516**, 321–323 (2014)

Famaey, B., Zhao, H.: Refining MOND interpolating function and TeVeS Lagrangian. Astrophys. J. Lett. **638**, 1–4 (2006)

Fattahi, A. et al.: The cold dark matter content of Galactic dwarf spheroidals: no cores, no failures, no problem. ArXiv eprints (2016) https://arxiv.org/abs/1607.06479. Accessed July 30, 2020

Jung, T., Beisbart, C.: Privileged, typical or not even that? our place in the world according to the Copernican and cosmological principles. J. Gen. Philos. Sci. **37**, 225–256 (2006)

Kragh, H.: Cosmology and controversy. Princeton University Press, Princeton (1996)

Kroupa, P.: The dark matter crisis: falsification of the current standard model of cosmology. Publ. Astron. Soc. Austral. **29**, 395–433 (2012)

Kuznetsov, I.: Printsip sootvetstviya v sovremennoy fizike i yego filosofskoye znacheniye. Gostekhizdat, Moscow (1948)

---

[25] As a concrete example, see the nominally Popperian cosmologist Ellis (2006). In the philosophy of cosmology, Butterfield (2014) is a good example of an approach to the discipline, where the tentativeness and continued lack of testability of the models is admitted, but the problem-shift seen as progressive nonetheless.

[26] I wish to thank John Antturi, Alexey Ilyin, David Miller, Zuzana Parusniková, David Merritt and Ilkka Niiniluoto for helpful discussions, suggestions and/or feedback regarding this chapter, or previous work that informs the content of this chapter. Naturally, none of the aforementioned individuals should be taken to endorse the views I have expressed here.

Lahav, O.: Observational tests for the cosmological principle and world models. In: Crittenden, R.G., Turok, N.G. (eds.) Structure Formation in the Universe. NATO Science Series C: (closed), vol. 565. Springer, Dordrecht (2001) https://doi.org/10.1007/978-94-010-0540-1_7

Lakatos, I.: Falsification and the methodology of scientific research programmes. In: Lakatos (1978), pp. 8–101 (1970)

Lakatos, I.: The methodology of scientific research programmes (philosophical papers volume I). In: Worrall, J., Currie, G. (eds.). Cambridge University Press, Cambridge (1978)

Lakatos, I.: Criticism and the methodology of scientific research programmes. Proc. Aristot. Soc. **69**, 149–186 (1969)

Lazutkina, A.: Theoretical terms and theoretical objects of astrophysics: a logico-methodological analysis. https://istina.msu.ru/diplomas/118631660/ (unpublished) Retrieved July 30, 2020

Lazutkina, A.: Theoretical terms of contemporary cosmology as intellectual artifacts. In: Franz, J. H., Berr, K. (eds.) Welt der Artefakte. Frank & Timme, Berlin (2017) https://arxiv.org/abs/1707.05235. Accessed July 30, 2020

Maartens, R.: Is the universe homogeneous? Phil. Trans. R. Soc. A. **369**, 5115–5137 (2016). https://doi.org/10.1098/rsta.2011.0289

MacIntyre, A.: After virtue. University of Notre Dame Press, Notre Dame (1981/2007)

Merritt, D.: Cosmology and convention. Stud. Hist. Philos. Sci. Part B: Stud. Hist. Philos. Mod. Phys. **57**, 41–52 (2017). https://doi.org/10.1016/j.shpsb.2016.12.002

Merritt, D.: A philosophical approach to MOND: assessing the Milgromian research program in cosmology. Cambridge University Press, Cambridge (2020) https://doi.org/10.1017/9781108610926

Milgrom, M.: A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. Astrophys. J. **270**, 365–370 (1983a)

Milgrom, M.: A modification of the Newtonian dynamics—implications for galaxies. Astrophys. J. **270**, 371–383 (1983b)

Milgrom, M.: A modification of the Newtonian dynamics—implications for galaxy systems. Astrophys. J. **270**, 384–389 (1983c)

Milgrom, M.: The MOND paradigm of modified dynamics. Scholarpedia **9**(6), 31410. (2014) http://www.scholarpedia.org/article/The_MOND_paradigm_of_modified_dynamics. Accessed July 30, 2020

Miller, D.: Popper's qualitative theory of verisimilitude. Br. J. Philos. Sci. **25**, 166–177 (1974b)

Miller, D.: On the comparison of false theories by their bases. Br. J. Philos. Sci. **25**, 178–188 (1974a)

Niiniluoto, I.: Truthlikeness. Reidel, Dordrecht (1987)

Niiniluoto, I.: Critical scientific realism. Oxford University Press, Oxford (1999)

Ntelis, P.: Probing cosmology with the homogeneity scale of the Universe through large scale structure surveys. Astroparticle and Cosmology Group, Physics Department, Paris Diderot University, Paris (2017)

Oddie, G.: Truthlikeness: the stanford encyclopedia of philosophy (Winter 2016 Edition). In: Zalta, E.N. (ed.) (2016) https://plato.stanford.edu/archives/win2016/entries/truthlikeness/. Accessed July 30, 2020

Popper, K.: Conjectures and refutations: the growth of scientific knowledge. Routledge, London (1963)

Popper, K.: Postscript to the logic of scientific discovery, vol. III, quantum theory and the schism in Physics. Rowman and Littlefield, Totowa (1982)

Popper, K.: Postscript to the logic of scientific discovery. In: Realism and the Aims of science, vol. I. Hutchinson, London (1983)

Ribeiro, C.: Karl Popper's conception of metaphysics and its problems. Principia **18**, 209–226 (2014)

Sarkar, P., Yadav, J., Pandey, B., Bharadwaj, S.: The scale of homogeneity of the galaxy distribution in SDSS DR6. Mon. Not. R. Astron. Soc.: Lett. **399**, L128–L131 (2009). https://doi.org/10.1111/j.1745-3933.2009.00738.x

Skordis, C., Złosnik, T.: A new relativistic theory for modified Newtonian dynamics. ArXiv e-prints (2020) https://arxiv.org/abs/2007.00082. Accessed July 30, 2020

Strigari, L.E., Bullock, J.S., Kaplinghat, M., Simon, J.D., Geha, M., Willman, B., Walker, M.G.: A common mass scale for satellite galaxies of the milky way. Nature **454**(7208), 1096–1097 (2008)

Sundell, P.: Beyond the cosmological principle. University of Turku, Faculty of Mathematics and Natural Sciences, Department of Physics and Astronomy, Turku (2016)

Tichý, P.: On Popper's definitions of verisimilitude. Br. J. Philos. Sci. **25**, 155–160 (1974)

Torretti, R.: Spacetime models for the world. Stud. Hist. Philos. Mod. Phys. **31**, 171–186 (2000)

Van der Waerden, B. (ed.): Sources of quantum mechanics. Dover Publications, New York (1967)

Voishvillo, E.K.: Printsip sootvetstviya kak forma razvitiya znaniy i ponyatiye otnositel'noy istiny. Kritika kontseptsii nesoizmerimosti smenyayushchikh drug druga teoriy. In: Markin, V.I., Ilyin, A.A., Zaitsev, D.V. (eds.) Logika i VEK. Sovremennyye tetradi Moskva, Moscow (2003)

Yang, A., Saslaw, W.: The galaxy counts-in-cells distribution from the SDSS. Astrophys. J. **729**, 123 (2011) http://dx.doi.org/10.1088/0004-637X/729/2/123

Anastasiia Lazutkina obtained her Bachelor's degree in Philosophy (Hons.) from Moscow State University. Her research interests include the philosophy and methodology of astrophysics and cosmology, logic, and formal methods in the philosophy of science. She is a published researcher and has appeared as an invited speaker at several international scientific and philosophical conferences. She currently studies mathematics and physics at LUT University, Finland. E-mail: a.k.lazutkina@gmail.com

Statistical Testing and Logic

Popper's Falsification and Corroboration from the Statistical Perspectives

Youngjo Lee and Yudi Pawitan

1 Introduction

Practical induction, including falsification and corroboration of propositions (scientific hypotheses, universal laws, social theories etc.), relies on information from limited data, hence must deal with uncertainty. Probability is the most recognized mathematical representation of uncertainty, yet Popper was emphatic in his rejection of inductive probability. Our goal is to bridge this gap. For the purpose of inductive reasoning, we can distinguish three concepts of probability. The first concept concerns the logical probability of a proposition being true as a measure of the degree of belief; this comes in two versions. The objective version (Keynes 1921) expresses a rational degree of belief, but it is never adopted by mathematicians or statisticians. The subjective version, proposed in a series of papers by Ramsey in 1920s (Ramsey 1931) and de Finetti in 1930s (de Finetti 1974), became the dominant version, and is later known as the Bayesian probability. The second probability concept relates to the long-run rate of observable and repeatable events (Von Mises 1928), mathematically formalized by Kolmogorov in 1933. And the third, Popper's propensity theory of probability defines it as the property of the generating mechanism—for instance, the coin toss. Frequentist statisticians interpret probability as having both the long-run and propensity properties. In orthodox statistical inference, the probability is associated with the P-value and the confidence intervals, routine quantities produced in virtually all scientific hypothesis testing. They are Popperian in the sense of being deductive quantities, but they are often interpreted inductively as evidence from observations.

Y. Lee
Department of Statistics, Seoul National University, Seoul, South Korea
e-mail: youngjo@snu.ac.kr

Y. Pawitan (✉)
Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
e-mail: yudi.pawitan@ki.se

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_7

In this paper we highlight the likelihood and confidence as alternative non-probabilistic measures of uncertainty that are Popper-compliant measures of corroboration. We shall discuss two well-known fallacies as specific illustrations. The conjunction fallacy has highlighted the limitations of probability-based reasoning as the rational basis of confirmation. The prosecutor's fallacy also arises from confusing probability and likelihood. To a large extent both fallacies vindicate Popper's objection to probability-based corroboration. We also discuss the results of a fascinating recent study that showed traces of likelihood-based reasoning in 15-month-old infants, indicating that such mode of reasoning is natural in human thinking.

In statistical terminology, anything unknown is called a parameter. It could the true speed of light; when we measure it, we only get an estimate with some margin of errors. The true lethality of the Covid-19 virus is an unknown parameter; all reported values are estimates that are uncertain because of statistical errors due to biased ascertainment and sampling variability. Statistical hypotheses are statements that can be re-expressed in terms of these underlying parameters. The classical likelihood was introduced by RA Fisher in 1921 to deal with unknown fixed parameters such as the speed of light or the viral lethality. There have been many attempts to extend the concept to accommodate random unknowns, but all failed (Lee et al. 2017). In 1996, Lee and Nelder introduced the extended likelihood approach for inferences of random unknowns. The confidence interval, perhaps the most important statistical quantity in scientific reports, has an unknown truth status. Whether or not a particular confidence interval covers the true parameter value is a random unknown. Theoretically it is a binary random variable, whose probability is represented by the confidence level. Confidence and likelihood have been fundamental statistical concepts with distinct technical interpretation and usage. We have shown recently (Pawitan and Lee 2020a) that confidence is in fact an extended likelihood, thus giving a much closer correspondence between the two concepts.

To illustrate, let G be a general proposition, such as "All ravens are black", and E be a particular proposition or an observation (evidence) such as "The raven in front of me is black." A deductive reasoning "G implies E" can attain complete confidence, given the basic premises—such as the axioms—are true. But, in practical induction, we want to establish G after observing E, which is logically possible but we know we may not attain complete confidence. We can use the logical probability to represent the deductive logic "G implies E" as

$$\mathrm{Prob(E|G)} = 1 \text{ and } \mathrm{Prob(not\ E|G)} = 0.$$

Probability extends classical binary logic, as the logical probability can be quantified as a number between 0 and 1 to capture a degree of belief. The extreme value 0 indicates impossibility (the proposition is false), and 1 certainty (the proposition is true). To use the likelihood for inductive inference, let θ be the probability of black ravens. Then G corresponds to $[\theta = 1]$. Given the observation (evidence) E, we define the likelihood as

$$\mathrm{L}(\theta) = \mathrm{Prob(E|}\,\theta).$$

We immediately see how the likelihood captures the information in E. If E is a white raven, then $L(\theta = 1) = 0$, meaning that, on observing a non-black raven, G has zero likelihood, hence falsified. Observing E being a black raven is certain when $\theta = 1$, so in this case $L(\theta = 1) = 1$. But it is also close to certain for θ near but less than 1, i.e. L(not G) ~ 1, so we cannot establish G based on observing a single black raven. But the evidence gathers strength if we go around the world observing 1000 ravens and find that they are all black. However, the likelihood function is never exactly zero for θ near but not exactly 1, indicating we are never certain.

The confidence approach is another practical implementation of Popperian views. Traditionally, given data E, construct confidence intervals with a certain probability of being true. For example, for the raven example: we want to establish G, which corresponds to $\theta = 1$. If we observe 10 ravens and all are black, the 95% confidence interval is 0.69 to 1.00 (Pawitan 2001, chap. 5). Thus, since 1.00 is in the interval, E does not falsify G that all ravens are black; it does support G, but the support is not that strong. Even when 31% of the ravens are non-black ($\theta = 0.69$), there is still a good chance to observe 10 black ravens in a sample of 10 ravens. Increasing the confidence say to 99% will only widen the interval 0.59 to 1.00. But increasing the sample size to 1000, the 95% confidence interval is now 0.996 to 1.000. So, there is a very strong support, but still not certainty, about G. Theoretically, we only get certainty with an infinite sample, but we can reach practical certainty. This simple example shows that likelihood and confidence can operationalize Popper's (1959) fallibilism for scientific development, where we can falsify, but cannot prove G; but it also clear that in practice we can build evidence for it. Traditional confidence procedure has a propensity interpretation: the 99% probability is the property of the procedure, not of the observed interval. So, it is fully Popperian. Recent developments in the confidence theory has brought the idea of confidence distribution, which is closely related to the Bayesian posterior probability. However, crucially, *confidence is still not a probability*, so it still conforms to the Popperian demand of non-probabilistic corroboration.

In summary, this simple example shows that likelihood and confidence can operationalize Popper's program for scientific development, where (i) we can falsify, but not prove G; (ii) they are not probability, so we can use them for non-probabilistic corroboration of hypotheses. Yet, there is a close logical connection between confidence and Bayesian subjective probability. In this sense, confidence bridges the gap between the inductive and Popperian non-inductive views. So we do not see a clear demarcation between these two views.

In an era of artificial intelligence, induction and learning from data becomes more crucial for drawing valid inferences. Originally, the goal of science was to prove propositions, to establish scientific theory based on observational data. However, the difficulty of such an inductive process has been recognized since the Greek and Roman periods. Hume (1748) argued that inductive reasoning cannot be justified rationally, because it presupposes that the future will resemble the past and the present. A resolution to the induction problem offered by Kant (1781) is to consider propositions as valid, absolutely a priori. Popper (1959) proposed falsification of propositions instead of proving them to be true. As Broad (1923) stated "induction is the glory of science but the scandal of philosophy”. However, it could be also a scandal of science if there is no way to confirm scientific theory with the complete confidence via induction. Jeffreys (1939) provided a way of confirming scientific theories by modifying Bayesian approach but requires an infinite evidence, which is not possible in practice.

Compared with scepticism, fallibilism does not reject the existence of knowledge. However, fallibilism recognize that there are not any reliable means of justifying knowledge as true or probable. Thus, justification is the misconception that knowledge can be genuine or reliable only if it is justified by some source or criterion (Deutsch 2011). Contemporary Bayesian confirmation theory has been developed to overcome various fallacy. The conjunction fallacy has highlighted the limitations of Bayesian probability-based reasoning as the rational basis of confirmation. We discuss the results of a fascinating recent study that showed traces of likelihood-based reasoning in 15-month-old infants, indicating that such mode of reasoning is natural in human thinking. In this article, we also show that via extended likelihood of Lee and Nelder (1996), one can obtain complete confidence (100% degree of belief) regarding a general proposition with finite evidence.

2 Probability-Based Reasoning for Inductive Inference

Popper was against probability-based induction, so it is instructive to see which particular aspects were anathema to him. Induction is a form of reasoning from particular examples to the general rule, in which one infers a proposition based on evidence (data or observations). However, establishing the truth of a general proposition is problematic, because it is always possible that a conflicting observation to occur. This problem is known as the induction problem. Originally, the goal of science was to confirm general propositions (scientific theories) such as “All ravens are black”, or to infer them from observational data. However, the difficulty of deriving such inductive logic has been recognized since the Greek and Roman periods.

As Cox (1946) noted, the logical probability represents the equivalence of ‘G implies E’ as

$$\mathrm{Prob(E|G)} = 1 (\text{and therefore } \mathrm{Prob(not\ E|G)} = 0)$$

with ‘not E implies not G’ as

$$\begin{aligned}\mathrm{Prob(G|not\ E)} &= \mathrm{Prob(not\ E|G)Prob(G)/Prob(not\ E)}\\ &= 0, \text{ and } \mathrm{Prob(not\ G|not\ E)} = 1,\end{aligned}$$

provided that the denominator is not zero. From this, we see clearly that one observation of a non-black raven can certainly falsify the general proposition. Popper (1959) saw this falsifiability of a proposition as a useful demarcation for scientific theory. Thus, he used deductive rule for scientific theory.

Via inductive rule, we want to establish G after observing E. The Bayes rule is

$$\mathrm{Prob(G|E)} = \mathrm{Prob(E|G)Prob(G)/Prob(E)} \tag{1}$$

provided Prob(E) > 0. Thus, if G implies E, we have

$$\mathrm{Prob(G|E)} = \mathrm{Prob(G)/Prob(E)} \geq \mathrm{Prob(G)},$$

with strict inequality if 0 < Prob(E) < 1, meaning that observing E increases the probability of G, unless Prob(G) = 0 or 1. Hence, a particular observation can corroborate, but it is not clear how much the evidence contributes to an increase of logical probability of the general proposition. Kant (1781) proposed a resolution to the induction problem, which involved considering the propositions as valid, absolutely a priori, i.e. Prob(G) = 1. He thought that Euclidean geometry is self-evidently true. In this case

$$\mathrm{Prob(G|E)} \geq \mathrm{Prob(G)} = 1, \text{ so } \mathrm{Prob(G|E)} = 1.$$

Thus, no evidence can increase of the logical probability of G. Broad (1923) stated that "induction is the glory of science but the scandal of philosophy" by showing that

$$0 = \mathrm{Prob(G|E)} \geq \mathrm{Prob(G)} = 0.$$

Popper (1959, Appendix vii) also explained why he believed that Prob(G) = 0 a priori and used it as argument against probability-based induction. If we know whether G is true or false a priori, its logical probability cannot be altered by future evidence. This seems to be a key aspect of the logical probability of induction that he found unacceptable. We show below that Prob(G) = 0 is actually not a necessary presumption.

2.1 The Logical Probability of Pascal

Throughout the article we define I(H) as the truth function of whether the proposition (or hypothesis) H is true or not, i.e.

$$\mathrm{I(H)} = 1 \text{ if H is true or } \mathrm{I(H)} = 0 \text{ if it is false.}$$

As the value of I(H) is unknown, let us treat it as an unobservable random variable. Then, we can represent logical probability as follows:

$$\mathrm{Prob(E)} = \mathrm{Prob[\ I(E)} = 1]\text{and}$$

$$\mathrm{Prob(G|E)} \;=\; \mathrm{Prob[\, I(G) = 1|I(E) = 1]} \;=\; \mathrm{Prob(G)/Prob(E)} \geq\; \mathrm{Prob(G)}.$$

Thus, Prob(G|E) = Prob(G), i.e. a new evidence cannot increase the logical probability of general proposition when Prob(G) is either 0 or 1; the two extreme cases of Kant and Popper.

Pascal's wager was published posthumously in *Pensées* ("Thoughts"). Given that reason alone cannot determine whether E is true or not, i.e. $T \equiv I(E) = 1$ or 0, Pascal concluded the uncertainty associated with this question can be expressed as probability Prob($T = 1$), treating T as an outcome from a coin toss but without seeing it. He further insisted even if we do not know the outcome of this coin toss, we must evaluate our actions based on the expectation of the consequence caused by our action. Pascal—one of the founders of probability theory—then treated his logical probability as Kolmogorov's probability. Pascal's argument of probability does not involve any data, so that he might be the first who used prior probability. He did not mention how to compute his prior probability. He also did not have any reason to distinguish his logical probability from Kolmogorov's because he lived in the seventeenth century, so unaware of twentieth century's mathematical probability. When we discuss the probabilities of Bayes (1763), Laplace (1814) and Fisher (1930), they also need not formulate the same probability as Kolmogorov's (1933) probability.

Pascal interpreted probability as logical probability of any proposition, such as Trump's impeachment. Ramsey and de Finetti interpreted it as a betting quotient and showed that, to be coherent, it should satisfy additive probability axioms. Otherwise someone can run a Dutch Book (arbitrage) on your betting. Suppose that you set your betting quotients as follows: (i) bet that the proposition E is true, risking to win $1 - \alpha$, and (ii) bet that E is not true, risking β to win $1 - \beta$. Then I will make *two bets* against you, on E and (not E) simultaneously, either as a player or as a bookie depending on $(\alpha + \beta)$ below. My expected winning from the two bets is

$$\begin{aligned}(1-\alpha)\mathrm{Prob(E)} - \alpha(1-\mathrm{Prob(E)}) + (1-\beta)\mathrm{Prob(not\ E)} - \beta(1-\mathrm{Prob(not\ E)})\\ = (1-\alpha-\beta)(\mathrm{Prob(E)} + \mathrm{Prob(not\ E)}) = 1-\alpha-\beta\end{aligned}$$

If $\alpha + \beta < 1$ I can always win by betting as player, and if $\alpha + \beta > 1$ I can always win by becoming a bookie. Thus, the betting quotients and β should satisfy probability laws such that $\alpha + \beta = 1$ and Prob(E) $= \alpha \geq 0$ and Prob(not E) $= \beta = 1 - \alpha \geq 0$. They call resulting logical probability subjective, because different people may have different betting quotients. According to Ramsey and de Finetti, we may know someone's personal probability via their betting behavior. Can people agree on their betting quotient? A question is whether there exists an objective logical probability. Objective Bayesians aim to have an objective logical probability by finding an ignorant prior probability, whereas we want to find it without presuming a prior.

The Bernoulli model was developed for observable binary random events such as coin tossing. In coin tossing, the true probability can be determined by long-run frequency (Von Mises 1928), whereas in logical probability the truth or falsity of proposition can never be observable, so that the long-run frequency is not available. Furthermore, expectation is for repeatable events not for a single event such as truthfulness of certain proposition. The interpretation of logical probability can be seen as an extension of propositional logic that enables reasoning with proposition whose truth or falsity is unknown. In this article, we show how recently developed likelihood theories helps to understand logical probability. We first briefly review developments of existing Bayesian approach in the last three centuries.

An inductive logic is based on the idea that the probability represents a logical relation between the proposition and the observations. Accordingly, a theory of induction should explain how one can ascertain that certain observations establish a degree of belief (logical probability) strong enough to confirm a given proposition. The sunrise problem is a quintessential example of the induction problem, which was first introduced by Laplace (1814). However, in Laplace's solution, a zero probability was assigned to the proposition that the sun will rise forever, regardless of the number of observations made. Therefore, it has often been stated that complete confidence regarding a general proposition can never be attained via induction. We explain why such an extreme view was formed. Lee (2020) shows that through induction, one can rationally gain complete confidence in general propositions via likelihood based procedure.

2.2 Bayesian and Fisher's Logical Probabilities

Pascal would first introduce logical probability but without the data and did not show how to compute it. It was Bayes (1763) who introduced Bayesian approach to set logical prior probability and update it based on the data. However, he might not have embraced the broad application scope now known as Bayesianism, which was in fact pioneered and popularized by Laplace (1814) as an inverse probability. Bayesianism has been applied to all types of propositions in scientific and other fields (Paulos 2011). Savage (1954) provided an axiomatic basis for the Bayesian probability as a subjective probability, whereas Jaynes (2003) provided as an objective probability. Whether it comes from objective or subjective Bayesian schools, Bayesians are common not to distinguish Kolmogorov's mathematical and logical probabilities because their axioms allow their logical probability satisfies laws of mathematical probability and require prior to form their logical probabilities. Fisher (1930) formed a logical probability without presuming a prior and axioms, so that his probability does not necessarily satisfy properties of Kolmogorov's mathematical probability even though he believed unfortunately so until his death.

2.3 Laplace's Solution to the Sunrise Problem

Using the sunrise problem, Laplace (1814) demonstrated how to compute such an actual logical probability based on the data. Let θ be the long-run frequency of sunrises, i.e., the sun rises on $100 \times \theta\%$ of days. Under the Bernoulli model, the general proposition G that "The sun rises forever" is equivalent to the hypothesis $\theta = 1$. The general proposition for which $\theta = 1$ is then a Popperian scientific theory because it can be falsified if a conflicting observation, i.e., one day of no sunrise, occurs. Based on finite observations until now, could it possible allow for complete confidence on $\theta = 1$?

To represent a description of the uncertainty about the true value of θ, prior probability should be assigned in Bayesian approach. Prior to the knowledge of any sunrise, suppose that one is completely ignorant of the value of θ. Laplace (1814) represented this prior ignorance about θ by means of a uniform prior on $\theta \in [0,1]$. This uniform prior was proposed by both Bayes (1763) and Laplace (1814) as Bayes-Laplace postulate of insufficient reason. Given the value of θ and no other information relevant to the question of whether the sun will rise tomorrow, Laplace computed the probability of the particular proposition E that "The sun will rise tomorrow." Laplace, based on a young-earth creationist reading of the Bible, inferred the number of days by considering that the universe was created approximately 6000 years ago. He computed the posterior, given $n = 6000 \times 365 = 102$ 2,190,000 days of consecutive sunrises,

$$\text{Prob}(\text{E}|\text{ n days of consecutive sunrises}) = 0.9999995;$$

see Lee (2020). This probability of this particular proposition, that is, the sun rising the next day given n days of consecutive sunrises, is eventually one as the number of observations n increases. However, this aspect is not sufficient to confirm the general proposition G that the sun rises forever. Broad (1918) showed that

$$\text{Prob}(\text{G}|\text{ n days of consecutive sunrises}) = 0 \text{ for all n}.$$

Hence there is no justification whatsoever for attaching even a moderate probability to a general proposition if the possible instances of the rule are many times more numerous than the instances already investigated for a more thorough discussion.

Jaynes (2003) argued that a beta prior density, Beta(α, β) with $\alpha > 0$ and $\beta > 0$, describes the state of knowledge that we have observed α successes and β failures prior to the experiment. The Bayes–Laplace uniform prior, Beta(1, 1), means that a trustworthy manufacturer sent you a coin with information that he/she observed one head and one tail in two trials before sending the coin, i.e. Prob(G) = 0. Even if you have an experiment with heads only for many trials, there is no way to attain complete confidence on heads only, unless you discard the manufacturer's information. Contemporary Jeffreys' prior (1939) and Bernardo's (1979) reference prior of objective Bayesian approach is Beta(1/2, 1/2), but Lee (2020) showed that under this objective Bayesian prior P(G| n days of consecutive sunrises) = 0. Thus, the Bayes-Laplace approach, even if it is derived from objective Bayesianism, cannot overcome the degree of skepticism raised by Hume (1748) because they presume Prob(G) = 0.

Let E be an event of "n consecutive days of sunrises," satisfying Prob(E|G) = 1. Laplace's (1814) solution shows that Prob(G|E) = 0 for any large n, because he presumed unfortunately Prob(G) = 0 a priori, which cannot be an ignorant prior. This leads to

$$0 = \text{Prob}(G|E) \geq \text{Prob}(G), \text{ so } \text{Prob}(G) = 0.$$

In his *Logic of Scientific Discovery,* Popper (1959) also explained why, he thinks, Prob(G) = 0, so that any evidence cannot alter the logical probability of general proposition Prob(G|E) = 0, which precludes probability-based induction.

2.4 Jeffreys's Resolution

That a general proposition cannot be confirmed via scientific induction based on the Bayes–Laplace formulation turns out to be because the choice of the prior had been wrong (implicitly presuming Prob(G) = 0). Jeffreys' (1939) resolution was another prior, which places a mass 1/2 on the general proposition $\theta = 1$ Prob(G) = 1/2 and a uniform prior on [0,1) with 1/2 weight. Let E be an event of n days of consecutive sunrises. This leads to

$$0 < \text{Prob}(G) = 1/2 < \text{Prob}(G|\text{one day of sunrises}) = 2/3$$

$$< \text{Prob}(G|\text{ two days of consecutive sunrises}) = 3/4$$

$$< \text{Prob}(G|E) = (n+1)/(n+2)$$

and thus, Prob(G|E) increases to one eventually as n increases (Lee 2020). Jeffreys' resolution produces an important innovation of the Bayesian hypothesis testing (Etz and Wagenmakers 2017). Senn (2009) considered Jeffreys' (1939) work as "a touch of genius," necessary to rescue the Laplacian formulation of induction. However, with Jeffreys's resolution, the scientific induction cannot attain complete confidence even in this era of big data, because such a process requires infinite evidence, i.e. P(G|E) = 1 only when the evidence is infinite. A key of Jefferys's resolution is to presume Prob(G) = 1/2 a priori. But there is another way to rescue the Laplacian formulation of induction without presumption of a prior.

2.5 Confirmation for General Proposition

Carnap's (1950) degree of confirmation of the general proposition G by the evidence E is

$$\mathrm{C(G,\ E)\ =\ Prob(G|E)\ -\ Prob(G)\ \leq 1 -\ Prob(G)\ =\ Prob(not\ G).}$$

Popper (1959) preferred 'corroboration' over 'confirmation', and 'testability' over 'confirmability,' since a theory is never confirmed to be true, i.e. Prob(G) = 0 for him. However, the term 'confirmation' has survived in the literature, so we also use it here to mean 'corroboration.' Since under Bayes-Laplace (Kant) formulation, Prob(G|E) = Prob(G) = 0 (Prob(G|E) = Prob(G) = 1) to give

$$\mathrm{C(G,\ E)\ =\ 0.}$$

In Carnap's inductive logic (1950), the degree of confirmation of every universal law is always zero. However, we see that it is because they presume Prob(G) = 0. Therefore, under Prob(G) = 0 a priori the universal law can never be accepted, but is not rejected until conflicting evidence appears. In Jeffreys' (1939) resolution with Prob(G) = 1/2,

$$\mathrm{C(G,\ E)\ =\ Prob(G|E) - Prob(G)\ = n/\{2(n+2)\}\ >\ 0,}$$

and thus the evidence E confirms the general theory G positively. But how we can justify the use of a prior, Prob(G) = 1/2. However, in the confidence resolution, although the prior Prob(G) is not assumed, complete confidence (confirmation) Prob(G|E) = 1 can be achieved with a finite evidence.

3 Confidence as an Alternative to Logical Probability

We see that Bayesian solution can give different conclusions, depending upon the choice of priors. Many writers have criticized the use of arbitrary priors. The question is whether we can form an objective logical probability without presupposing a prior. In 1930 Fisher showed that it is possible and called the idea "fiducial probability," which has largely been abandoned in practical statistics, but it did lead to the "confidence" concept. Confidence interval is one of the most widely used statistical inference tools in practice.

For example, in polls before election, the 95% confidence interval of the true voting rate θ_0 of a certain candidate is reported. Based on poll data, suppose that the $\alpha \times 100\%$ confidence interval [L, U] is reported. Let

$$\mathrm{T}\ \equiv\ \mathrm{I(L}\ \leq \theta_0 \leq\ \mathrm{U),}$$

where $T = 1$ if the interval covers the truth, and 0 otherwise. Thus, T (as a function of data) is a binary random variable with probability

$$\text{Prob}(T = 1) = \alpha.$$

When it *refers to the observed interval*, α is called the confidence, rather than probability, of the interval because it is not a proper Kolmogorov probability (Schweder and Hjort 2016). Let E be the proposition that the specific interval [L, U] covers the true parameter θ_0 and

$$T = I(E).$$

We shall see that the confidence is indeed a way of computing Pascal's logical probability of proposition and an alternative to Bayesian probability without assuming a prior. The confidence statement of the interval, the true value of θ is contained in the interval, is sample dependent proposition, and we attain complete confidence when $\alpha = 1$.

Fisher's (1930) classical likelihood is for inferences about fixed unknowns. Lee and Nelder (1996) extended likelihood inferences to random unknowns such as T. Pawitan and Lee (2020a) showed that the confidence is the extended likelihood of Lee and Nelder (1996). For an observed confidence interval [l, u], the value $T = t = I(l \leq \theta_0 \leq u)$ is realized but still unknown single event, because the true parameter value θ_0 is unknown, giving

$$L(t = 1) \equiv \text{Prob}(T = 1) = \alpha \tag{2}$$

Here $L(t = 1)$ is the extended likelihood of single (realized) event; it is equal to the confidence. The confidence interval is justified theoretically in terms of its coverage probability, which is given a Popperian propensity interpretation as belonging to the procedure (Gillies 2000). But in (2) we address the epistemic question if there is a probabilistic way to state our sense of uncertainty in an observed confidence interval. In coin tossing, we can compute the long-run frequency as a true probability. However, in the confidence concept the realized value t is unobservable, so its long run frequency is not meaningful. Instead, frequentists use coverage probability in hypothetical repetitions of constructing confidence intervals as a thought experiment. If we construct confidence interval 100 times repeatedly from the same experiments, 100 x α of them will cover true value of θ. The coverage probability is a long-run rate of the coverage of the confidence interval in hypothetical repetitions. Thus, the confidence concept is a bridge between the Kolmogorov and logical probabilities.

The P-value has been widely used for scientific inferences. Let X be a sufficient statistics for θ. Fisher (1930) derived the fiducial probability of θ. Define the right-side P-value function

$$C(x, \theta) = \text{Prob}(X \geq x \mid \theta).$$

Given X = x, as a function of θ, C(x, −∞) = 0 and C(x, ∞) = 1 and C(x, θ) is a strictly increasing function of θ. Thus, C(x, θ) behaves as if it is the cumulative distribution of θ. This leads to the fiducial probability for θ

$$c(x,\ \theta)\ =\ dC(x,\ \theta)/d\theta,$$

which is derived without presupposing a prior. Schweder and Hjort (2016) called it the confidence density. Pawitan and Lee (2020a) showed that this sample-dependent confidence is indeed an extended likelihood with updating rule

$$c[(x1, x2),\ \theta] \propto c(x1, \theta)L(\theta;\ x2),$$

where c[(x1,x2), θ] is the confidence based on the combined data (x1,x2), c(x1,θ) is that based on the data x1, and L(θ; x2) = Prob(X2 = x2| θ) is the likelihood based on the data x2. This leads to

$$c(x,\ \theta) \propto c0(\theta)L(\theta;\ x),$$

where c0(θ) ∝ c(x, θ)/L(θ; x) is the induced prior confidence without the data, and L(θ; x) the likelihood.

The confidence density c(x, θ) and induced prior c0(θ) correspond to the Bayesian posterior Prob(θ| x) and prior Prob(θ), respectively. Thus, the confidence can be obtained by using the Bayes rule (1) under the induced prior confidence. So, confidence is the frequentist alternative to Bayesian logical probability. However, confidence is derived without presuming a prior, whereas Bayesian posterior is based on the prior. But confidence and therefore fiducial probability is not necessarily a Kolmogorov probability, so that we use c(x, θ) to represent the confidence even though it plays the same role as the Bayesian logical probability. In the Bayesian approach as long as a prior is proper Kolmogorov probability, the posterior is always proper. However, induced prior c0(θ) is often improper, the resulting confidence may not be a proper probability. Just as a Bayesian posterior contains a wealth of information for any type of Bayesian inference, a confidence density contains a wealth of information for constructing almost all types of frequentist inferences on fixed parameter θ (Xie and Singh 2013). For notational convenience, we shall sometimes use Bayesian logical (posterior) probability Prob(θ| x) under a prior c0(θ) for the confidence density c(x, θ) in the confidence approach.

To summarize, confidence and likelihood are fundamental statistical concepts, currently known to have distinct technical interpretation and usage. Confidence is a meaningful concept of uncertainty within the context of confidence-interval procedure, while likelihood has been used predominantly as a tool for statistical modelling and inference given observed data (Pawitan 2001; Lee et al. 2017). Pawitan and Lee (2020a) showed that confidence is an extended likelihood, thus giving a much closer correspondence between the two concepts. This result gives the confidence concept an external meaning outside the confidence-interval context, and the extended likelihood theory gives a clear way to update or combine confidence information. On the other hand, the confidence connection gives the extended likelihood direct access to the frequentist confidence interpretation, an objective certification not directly available to the classical likelihood. This implies that inferences from the extended likelihood have the same logical status as confidence interpretations, thus simplifying the terminology in the inference of random parameters.

3.1 Confidence Resolution of Induction Problem

Let E be the proposition "The sun rises tomorrow" and G be that "it rises forever." Lee (2020) derived a confidence density for the sunrise problem using Pawitan's (2001, chap. 5) right-side P-value, leading to a logical probability

$$\text{Prob}(\text{G}|\text{ n days of consecutive sunrises}) = 1$$

so that

$$\text{Prob}(\text{E}|\text{ n days of consecutive sunrises}) = 1$$

This allows the realization of complete confidence even with $n = 1$. Furthermore,

$$\text{Prob}(\text{G}|\text{ no sunrise at least in one day}) = 0$$

Russell (1912) illustrated induction problem, "Domestic animals expect food when they see the person who usually feeds them. We know that all these rather crude expectations of uniformity are liable to be misleading. The man who has fed the chicken everyday throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken." Regardless of the number of observations, Hume (1748) would even argue that we cannot claim it is"more probable", since this still requires the assumption that the past predicts the future. Let G be a general proposition that a specific event E occurs always. Let $Xi = 1$ if E occurs at the ith independent observation or experiment and $= 0$ otherwise. The long-run frequency of $Xi = 1$ is θ. Provided

$$Tn = X1 + \ldots + Xn = n,$$

we can claim the uniformity of nature that the event E occurs always with complete confidence. Both Russell and Hume presume non-uniformity $\text{Prob(G)} = 0$.

In response to the skepticism raised by Hume (1748), Kant (1781) proposed the consideration of the general proposition as absolutely valid $\text{Prob(G)} = 1$, a priori, which is otherwise drawn from the dubious inferential inductions. In contrast Bayes (1763) and Laplace (1814) presumed a priori that the general proposition is false $\text{Prob(G)} = 0$. Thus, Kant's proposal is consistent only if the general proposition is true, whereas the Bayes–Laplace rule is consistent only if the general proposition is false. It is not necessary a priori to presume Prob(G) = 0 or 1. Jeffreys (1939) presumed Prob(G) = 1/2 to rescue the Laplacian formulation of induction. Now we discuss why the confidence approach provides a resolution of induction problem. Lee (2020) demonstrated that the confidence leads to two potential induced priors, specifically, Beta(0,1) and Beta(1,0). Although these priors are not proper probability, having an infinity measure, they allow a reasonable interpretation. For example, the Beta(1,0) prior indicates that only one success is observed a priori. Thus, if we observe all successes, it is legitimate to attain 100% confidence on $\theta = 1$. However, even if we observe all the failures, we can never attain 100% confidence on $\theta = 0$ because of the success a priori. The Beta(0,1) prior exhibits the contrasting property.

Through deduction, one can achieve complete confidence regarding a particular proposition

$$\mathrm{Prob}(E|G) = 1,$$

provided that the general proposition G is true, Prob(G) = 1. Through induction, we can have $0 \leq \mathrm{Prob}(G|E) \leq 1$. Under Bayes-Laplace (Kant) formulation Prob(G|E) = Prob(G) = 0 (Prob(G|E) = Prob(G) = 1) for any evidence E because of presumption Prob(G) = 0 (Prob(G) = 1) a priori, whereas under Jeffrey's resolution $P(G|E) > 0$ because he presume Prob(G) = 1/2 but cannot reach one (complete confidence) with finite evidence. However, the confidence resolution implies surprisingly that it is legitimate to claim P(G|E) = 1, i.e. one can attain complete confidence regarding the general proposition in finite samples. Confidence approach interprets such a complete confidence as a consistent sample dependent frequentist estimator of the unknown logical probability P(U = 1) = P(I(G) = 1). The estimator becomes more accurate as evidence grows. Lee (2020) showed its theoretical consistency further.

To confirm the validity of the general relativity theory, the observational evidence of light bending was obtained in 1919 and the astrophysical measurement of the gravitational redshift was obtained in 1925. Thus, a new theory was confirmed based on a few observations. Then, our resolution shows that it is legitimate to predict the future uniformly with complete confidence unless the general relativity theory stops to hold in the future. Such an inductive reasoning is theoretically consistent and therefore rational. Via induction based on finite data, we can complete confidence that the sun rises forever. (Of course, in physics, the sun runs out of energy, and the solar system vanishes eventually, but here we are discussing only our logical-mathematical confidence given some evidence). To establish the general proposition from the particular instances by means of induction, scientists do not need to review all the instances but to establish a scientific theory pertaining to the generation of the instances. If one drops an apple, one can be sure that it will fall unless the Newtonian laws suddenly stops to hold. Indeed, it is induction, as we have seen, to allow such uniformity, so that it is the glory of both science and philosophy.

4 Extended Likelihood as Objective Logical Probability

The likelihood is an uncontroversial technical element that is acceptable to all schools of statistics, but its direct use for inference is controversial (Pawitan 2001, section 2.6). Fisher (1973) recognized the two logical levels of uncertainty, one captured by probability and the other by his classical likelihood, correspondingly two levels of rational thinking. They are not meant to be in competition with each other, as classical likelihood is weaker than logical probability. For example, unlike Bayesian logical probability, the classical likelihood values do not allow an objective frequency-based calibration. Fisher (1930) proposed "fiducial probability" as an alternative to Bayesian logical probability. However, controversies arise as fiducial probability is not necessarily Kolmogorov probability, so that it has been abandoned. Recently, in statistical literature it is appearing as confidence. We claim that the confidence is objective because it can be obtained without assuming a subjective prior. Pawitan and Lee (2020a) showed the confidence is indeed an extended likelihood for unobserved random variable T

$$\mathrm{Prob}(\mathrm{T} = \mathrm{I}(\mathrm{E}) = 1) \tag{4}$$

which can be viewed as a betting quotient of an event (or proposition) E. Ramsey (1931) and de Finetti (1974) proved that it is coherent as long as it satisfies probability laws. But their definition of logical probability (4) is subjective because people with the same data are allowed to have different logical probabilities. With confidence approach, the confidence (4) is an objective extended-likelihood value, which does not depend upon subjective priors. It can be an objective betting quotient unless there is a relevant subset (Fisher 1958).

4.1 *Postulate of Ignorance*

In the sunrise problem Laplace adopted the principle of insufficient reason to justify the use of the uniform prior. We see that it implicitly presumes $\mathrm{Prob}(\mathrm{G}) = 0$, so that it cannot be an ignorant prior. Fisher (1958) noted that the postulate of ignorance is very important in developing inductive methods. However, he refused to make any axiomatic prior probability. Thus, his fiducial probability (and therefore confidence) does not necessarily satisfy properties of Kolmogorov's probability (Schweder and Hjort 2016), whereas Bayesian logical probability does if the prior is a proper probability.

According to Gödel (1931), even in a mathematical deductive system, there always exists a proposition G that can be neither proved nor disproved. Thus, mathematics itself also cannot avoid uncertainty. Turing (1936) reformulated Gödel's 1931 results, replacing Gödel's universal arithmetic-based formal language with a simple hypothetical devices known as Turing machine, which is capable of performing any conceivable computation. Turing machine is realized as modern computer. To rephrase Gödel's 1931 problem: can a computer determine whether an arbitrary proposition can be proven or not? Turing (1936) showed that the answer is no. He proved that it is not possible to decide whether a Turing machine will ever halt to return the answer. Uncertainty is also unavoidable in many different ways even in computing. In summary, there is a mathematical proposition G which cannot be proven even if they are true. Likewise, a computer cannot tell us whether the proposition G is solvable in finite time or not. Thus, even in mathematics and computer sciences there is always a proposition G, whose truthfulness is unknown:

$$\mathrm{T} = \mathrm{I(G)} = 1(\text{G is true}) \text{ or } 0\ (\text{G is not true})$$

Thus, even though Prob(G) is either 0 or 1, but it is not possible to know Prob(G) a priori. Human may not perceive the true general laws, but it does not mean Prob(G) = 0. Furthermore, to presume Prob(G) = 0 a priori does not imply ignorance neither. However, to postulate an ignorance, namely the principle of insufficient reason, Bayes and Laplace used the uniform prior, which turns out to presume Prob(G) = 0. We also see Jeffreys's (1939) ignorant prior and Bernardo's (1979) of objective Bayesian schools reference prior presume Prob(G) = 0. This is as a strong presumption as Kant's Prob(G) = 1. In his rejection of inductive reasoning, Popper also presumed Prob(G) = 0, so that in consequence he believed the falsification is the only way to conduct scientific inference. In this formulation Popper's view on inductive reasoning was as extreme as Kant's. We can confirm or falsify general proposition based on evidence, which can be turned out to be wrong later according to future evidence. But that is what our human can do confidently (corroborate) in building our knowledge.

In the sixteenth century, Michel de Montaigne was most famously known for his skeptical remark, "What do I know?" He was the one who thought deeply about ignorance and concluded that it could not be ended with period. If we are saying we do not know something, then it cannot be an ignorance because we know what is unknown. The best way to represent the ignorance would be to do nothing about it, following Wittgenstein (1921) "What we cannot speak about we must pass over in silence." To represent ignorance by specifying Prob(G) = 0 is not ignorance at all, i.e. it is knowledge that G is known to be false a priori.

Deductive logic is based on the knowledge or assumption that certain general propositions or axioms are true. Thus, its conclusions, any derivable statements, are true without any uncertainty. However, inductive reasoning mainly concerns handling of uncertainties, caused by lack of information (ignorance) or limited data. For inductive reasoning, Bayesian school also uses deductive logic by presuming the prior on Prob(G) (like making an axiom in mathematics). Whereas, with the confidence or extended likelihood approach, we do nothing on what we do not know; this, we believe, is the most important requirement in developing inductive reasoning. Popper was against the Bayesian logical probability approach, so the non-Bayesian approach here can help to realize some of his visions.

5 Confirmation Problems

We now discuss the difficulties arising in confirming propositions via two well-known fallacies—the conjunction and prosecutor's fallacies. They highlight two distinct modes of reasoning, one captured by logical probability and the other by (classical) likelihood. Our seemingly irrational behavior is due to a decision making based on the likelihood. Thus, from the likelihood perspective, we are still behaving rationally. Recognizing these two modes may lead to better understanding and assessment of our decisions. The difficulties vindicate Popper's criticism of probability-based induction. Also known as Linda problem, the conjunction fallacy originated from Tversky and Kahneman (1983):

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable?

H1 Linda is a bank teller.

H2 Linda is a bank teller and is active in the feminist movement.

Because H2 is a conjunction of two events (hypotheses), it always has lower probability than H1. Yet, from Kahneman (2011, page 158): "About 85 to 90% of undergraduates at several major universities chose the second option, contrary to logic. Remarkably, the *sinners* [our emphasis] seemed to have no shame. When I asked my large undergraduate class in some indignation, 'Do you realize that you have violated an elementary logical rule?' someone in the back row shouted, 'So what?'" After seeing the results of their empirical studies he wrote: "I quickly called Amos [Tversky] in great excitement to tell him what we had found: we had pitted logic against representativeness, and representativeness had won!"

Using probability-based reasoning for Linda problem we are seemingly forced to prefer H1 over H2 *regardless of the data.* This actually feels unnatural: in science it is more reasonable to assume that scientists will formulate and test the strongest hypothesis that is supported by the data, not the safest. This actually corresponds to a seemingly paradoxical Popperian view that, among competing hypotheses, one should in fact adopt the least probable hypothesis that is supported by data. The qualifier "supported by data" in practice of course requires a statistical test. The safest hypothesis can be the blandest, the one with the highest probability of being correct, but has the weakest power to explain the data. 'A feminist bank teller' is 'a bank teller' for sure, but the feminist element makes it a more interesting hypothesis with more explanatory power than the bland 'bank teller'.

We could also add other boring hypotheses such as 'H3: Linda is a woman' or even 'H4: Linda is human', which would have higher probabilities than 'H1: Linda is a bank teller'. Is it 'rational' to prefer H4 when there is enough information pointing to H2? The preference of H2 over H1 is an indication that—for the people making such judgement—there is enough information pointing to H2. On the other hand, it is not reasonable either to choose 'H5: Linda is a widowed feminist bank teller' as there is nothing in the data supporting the widowed status. So, rationally, the best hypothesis is the strongest hypothesis that is supported by the data; this relates to a notion of optimality, and much hypothesis tests based on the likelihood ratios have been developed to establish the optimality of likelihood-based inference, e.g. Neyman and Pearson (1933).

5.1 Bayesian Reasoning in the Conjunction Fallacy

The conjunction fallacy has been used as an example of a defect in human reasoning. Despite extensive inquiry, however, the attempt to provide a satisfactory account of the phenomenon has proved challenging. Bayesian confirmation theory has been developed. Inductive logic may be seen as the study of how data (evidence) affect the probability of a proposition H. From Laplace (1814), the posterior Prob(H|data) is considered as an appropriate formalization of the basic inductive logical relationship between evidence and proposition. However, this could lead to counterintuitive consequences and conceptual contradictions (Popper 1959). There is a fundamental distinction between the notions of logical probability (firmness) and its increase in a proposition H in the light of evidence. Thus, the posterior (logical probability) could be taken as accounting for the former concept, but not the latter (Carnap 1950). In fact, the degrees of belief on H may increase as an effect of evidence and still remain relatively low (for example, because the disease of interest is very rare). The term"-confirmation" has been used in the epistemology and philosophy of science whenever the observational data (evidence) support scientific proposition, meaning in terms of Carnap's increase in firmness brought by data to H. Many Bayesian confirmation measures have been proposed. As an example we consider Carnap's (1950) degree of confirmation of proposition H by the data (evidence)

$$\mathrm{C(H,\ Data)} = \mathrm{Prob(H|Data)} - \mathrm{Prob(H)} \tag{3}$$

which can be positive unless Prob(H) is either 0 or 1. Crupi et al. (2008) derived somewhat complicated conditions under which all confirmation measures satisfy

$$\mathrm{C(H2,\ Data)} \geq \mathrm{C(H1,\ Data)}$$

Thus, increase of the probability of H2 by the data can be greater than that of H1. However, a difficulty in Bayesian confirmation is again how to choose the prior, so that it seems arbitrary and complicated conditions are necessary for confirmation.

5.2 Likelihood Reasoning in the Conjunction Fallacy

The literature on the conjunction fallacy unfortunately does not make a distinction between 'probability' and 'likelihood'. Consider the first part of the description of Linda problem (Linda's characteristics) up to the question as 'Data', and the two statements about her as hypotheses H1 and H2. Then the assessment of H1 and H2 can be either probability-based or likelihood-based. Mixing them up generates the apparent fallacy and confusion.

The likelihood-based reasoning is based on comparing the classical likelihoods of Fisher (1921)

$$\mathrm{L1} = \mathrm{L(H1)} = \mathrm{L(Linda\ is\ a\ bank\ teller)} \equiv \mathrm{Prob(Data|H1)}$$

$$\mathrm{L2} = \mathrm{L(H2)} = \mathrm{L(Linda\ is\ a\ feminist\ bank\ teller)} \equiv \mathrm{Prob(Data|H2)},$$

while the probability-based reasoning is based on comparing the logical probabilities

$$\mathrm{P1} = \mathrm{Prob(Linda\ is\ a\ bank\ teller|Data)} = \mathrm{Prob(H1|Data)}$$

$$\mathrm{P2} = \mathrm{Prob(Linda\ is\ a\ feminist\ bank\ teller|Data)} = \mathrm{Prob(H2|Data)}.$$

Now, as probabilities, it is always the case that $\mathrm{P2} \le \mathrm{P1}$. But as likelihoods, there is no guarantee at all that $\mathrm{L2} \le \mathrm{L1}$, because likelihood is not a probability of hypothesis.

In Linda problem it is possible that, *when given the description*, the study participants are intuitively making their judgement between feminist vs non-feminist alternatives, thus actually preferring H2 over an unstated but more natural competing hypothesis of non-feminist bank teller. Is the likelihood-based reasoning still consistent with the choice of H2 over H1? Thus, suppose that we have the complementary hypotheses

H2 Linda is a feminist bank teller vs

H3 Linda is a non-feminist bank teller,

and further assume that Linda is a typical female bank teller, i.e. not specially selected, so we can logically compute the necessary probabilities. In likelihood terms, the preference of H2 over H3 is the judgement that

$$\mathrm{L2} = \mathrm{Prob(Data|H2)} \ge \mathrm{L3} = \mathrm{Prob(Data|H3)}$$

First note that $\mathrm{H1} = \{\mathrm{H2\ or\ H3}\}$, i.e. a bank teller is either a feminist or a non-feminist. Among the female bank tellers, let us denote the proportion of feminists as p and the proportion of non-feminists $(1 - \mathrm{p})$. Then, with some probability calculations we have

$$\begin{aligned} \mathrm{L1} \equiv \mathrm{L(H1)} &= \mathrm{Prob(Data|H1)} \\ &= \mathrm{Prob(Data|H2\ or\ H3)} \\ &= \mathrm{p} \times \mathrm{L2} + (1 - \mathrm{p}) \times \mathrm{L3} \\ &= \mathrm{L2} - (\mathrm{L2} - \mathrm{L3}) \times (1 - \mathrm{p}) \\ &\leq \mathrm{L2} \end{aligned}$$

because the term (L2 − L3) × (1 − p) ≥ 0 for any value of p between 0 and 1. Hence the composite hypothesis H1 has a lower likelihood than the constituent likelihood H2. So, within the likelihood framework, the order of preference between H2 and H3 is consistent with the order between H2 and H1.

This is in stark contrast to logical probability-based reasoning, since we always have Prob(H1|Data) ≥ Prob(H2|Data) regardless of the ordering of Prob(H2|Data) vs Prob(H3|Data). In fact, by taking the 'Data' into account, a great majority of the undergraduates and the homunculus are thinking the opposite: implicitly making the judgement that L2 ≥ L1, hence preferring the second hypothesis. In likelihood approach, likelihood ratio L2/L1 ≥ 1 is used to select H2 between hypotheses H1 and H2.

Which reasoning is better? This is not a simple question. Mathematically it depends on how 'Linda' comes into the picture. If she is randomly selected from the population (which determines the sampling probability of Linda), the extended likelihood is defined (Lee et al. 2017). Then, probability(confidence)-based reasoning is mathematically guaranteed to be better, in the sense that it would produce less error. In such cases, the confidence-based reasoning can be also justified under the extended likelihood framework (Lee et al. 2017). If sampling distribution is proper, the confidence is proper probability.

However, the problem description does not make any explicit statement how Linda was selected. Without the random selection, then in principle there is no definite answer; e.g., in the study Linda could be specially selected from among the feminists, in which case H2 is correct. There are many scientific studies that do not rely on random samples. For example: (i) in clinical trials we randomize subjects into study groups; (ii) in epidemiologic studies subjects are often selected based on their outcomes status, resulting in non-random selection. The (classical) likelihood-based reasoning presumes that we know nothing about how Linda comes to the picture, so the likelihoods are the only available *objective quantities* for inference. By preferring H2, the undergraduates are making this stance implicitly. Is that irrational?

As we have stated above, the largest section of scientific statistical data analysis today is based on the classical likelihood. That is, probabilities about the states of nature are rarely included in the analyses, so the analysis is closer in spirit to the classical likelihood-based reasoning above. It is possible to state the problem more carefully so that the logical probability (Bayesian posterior or frequentist confidence) is a better metric for decision, for example by making explicit that Linda was chosen randomly from among 100 women that fit the description, but elaborating on such a situation is not our purpose here. We simply want to provide an explanation of the conjunction fallacy, which is that many people—including the sophisticated undergraduates and homunculus—appear to use classical likelihood-based reasoning in daily life when sampling probability of Linda is uncertain. Hence the 'conjunction fallacy' is not a fallacy, but the result of a mathematically valid likelihood-based reasoning.

6 Prosecutor's Fallacy

The so-called prosecutor's fallacy can also be explained as a confusion between probability and likelihood-based reasoning. The application of statistical inference in court has been the subject of serious discussion and debates, especially after the emergence of DNA profiling as part of evidence. The logic of legal concepts such as 'presumed innocence' or 'guilt beyond reasonable doubt' has direct statistical connotations, so the principles apply more generally to any assessment of evidence. As discussed in Gardner-Medwin (2005), three key propositions at issue:

A: the facts or evidence could have arisen if the defendant is guilty
B: the facts or evidence could have arisen if the defendant is innocent
C: the defendant is guilty.

Clearly A and not-B together would imply C, but C does not imply not-B. The latter is obvious if the evidence is weak, i.e. could easily have been found among innocent people. Thus strong beliefs in A and non-B together is a more stringent requirement than a belief in C alone. In fact, for expert witnesses, the presumed-innocence requirement may preclude the assessment of C. Since the categorical truth of these statements is in reality rarely available, the prosecution may have to present extremely small probabilities to establish guilt beyond reasonable doubt.

Those probabilities of A and B are in fact likelihoods of guilt:

$$\mathrm{L1} = \mathrm{L(Defendant\ innocent)} = \mathrm{Prob(Evidence|Defendant\ innocent)}$$

$$\mathrm{L2} = \mathrm{L(Defendant\ guilty)} = \mathrm{Prob(Evidence|Defendant\ guilty)}$$

while the probabilities of guilt are

$$\mathrm{P1} = \mathrm{Prob(Defendant\ innocent|Evidence)}$$

$$\mathrm{P2} = \mathrm{Prob(Defendant\ guilty|Evidence)}.$$

Thus in a typical court proceeding, what is computed is the likelihood L1. The so-called prosecutor's fallacy is to misrepresent L1 as P1. Putting aside any technical issues in its computation, suppose L1 is very low, say 10^{-8}. And suppose further that the probability of DNA matching is one if the defendant is guilty, i.e., L2 =

Prob(Evidence|Defendant guilty) = 1. So the likelihood-based reasoning leads to a likelihood ratio

$$\text{L2/L1} \;= 10^8,$$

which provides the joint assessment of (A and non-B) propositions. In using L1 and L2 directly without prior probabilities of guilt, the prosecutor is relying on a *valid likelihood-based reasoning. It is not a fallacy.*

One may of course argue that probability-based reasoning based on P2/P1 is better, but this will require the establishment of prior probabilities of guilt that can be agreed by all parties. One can easily imagine the contentious arguments on settling the prior probabilities; for example, is it reasonable to presume that the defendant is a random sample from the general population? How do we abide by the presumed innocence requirement? However, again our purpose here simply to point out that the prosecutor's argument in fact does not have to rely on the logical probabilities (confidences), but on a valid (classical) likelihood-based reasoning, hence avoiding the fallacy. Unfortunately, in layman language, 'probability' and 'likelihood' are interchangeable as expressions of uncertainty, thus confusing the two valid modes of reasoning and making it impossible for the prosecutor to avoid the fallacy.

7 Traces of Likelihood-Based Learning in Infants

Which is more natural: probability-based or likelihood-based reasoning? Gweon et al. (2010) presented a fascinating series of experiments on inductive learning by infants—average age 15 months old—as evidence that likelihood-based learning is perhaps hard-wired in our brains. The story is delightfully told in Laura Schulz's Ted Talk: How do babies' "logical minds" work? https://www.ted.com/talks/laura_schulz_the_surprisingly_logical_minds_of_babies.

Here we only highlight their key experiment: the babies were presented with *3 squeaky blue balls* taken from a large opaque box. The box-wall facing the babies had a clearly visible picture indicating its content. Two scenarios were performed:

Scenario 1: the picture showed mostly blue balls and some yellow balls;
Scenario 2: the picture showed mostly yellow balls with some blue balls.

(The ratio is 3:1 in each case, but as far as the babies were concerned we suppose the exact number did not really matter.) The blue balls were taken one at a time, each time shown to the babies that they squeaked. Then babies were given a *single yellow ball*; the question is would they attempt to squeak it? What reasoning or inference method do they use?

First we can agree that the only Data available to the babies are {3 squeaky blue balls}. To use the probability-based reasoning the babies would have to come with the Bayesian posterior probability

$$\text{Prob}(\text{yellow balls are squeaky}|\text{Data}),$$

which would of course require the prior probability before seeing the data and the use of Bayes's formula to compute the posterior. Presumably, if the baby judges the probability to be high enough, then they will try to squeak the given yellow ball. However, the necessary calculation looks too difficult for most babies we know. The likelihood-based reasoning would require the babies to assess

$$\text{L}(\text{yellow balls are squeaky}) = \text{Prob}(\text{Data}|\text{yellow balls are squeaky}).$$

This is perhaps not obvious either, as there is never any direct evidence of the squeakiness of yellow balls; so, the inference must somehow come from an inductive generalization.

Gweon et al. (2010) described a model involving 4 hypotheses leading to predictions based on likelihood reasoning, but here we shall construct a simpler thought process. On seeing 3 squeaky blue balls, the babies were implicitly assessing these 2 hypotheses:

H1 the sample is randomly selected from all the balls

H2 the sample is not random, but selectively taken only from squeaky blue balls

Furthermore, with inductive generalization, when a sample was judged random then the properties of the sample would generalize; for instance, here, squeakiness then applies to all balls, hence to the yellow balls. And vice versa, when a sample was not random, then the properties would not generalize.

On observing 3 squeaky blue balls, the likelihoods of the hypotheses are now computable. For Scenario 1 (mostly blue balls):

$$\text{L1} = \text{Prob}(\text{Data}|\text{H1}) \sim \text{high}$$

$$\text{L2} = \text{Prob}(\text{Data}|\text{H2}) \sim \text{high}.$$

Presumably the babies did not use the exact values, but used only visual clues ($\text{L1}/\text{L2} \sim 1$) to conclude that there was no reason to reject the random sampling hypothesis H1. So the squeaky property generalized to yellow balls, and the babies were predicted to squeak the yellow balls. On the other hand, for Scenario 2 (mostly yellow balls):

$$\text{L1} = \text{Prob}(\text{Data}|\text{H1}) \sim \text{low}$$

$$\text{L2} = \text{Prob}(\text{Data}|\text{H2}) \sim \text{high}.$$

Again using only visual clues ($\text{L1}/\text{L2} < < 1$) to judge the hypothesis, babies would reject the random sampling hypothesis H1, hence not generalize the squeaky property. So they were predicted not to squeak the yellow balls. The results confirmed these predictions, or we could say the likelihood-based reasoning explains the experimental results, providing evidence of elementary use of likelihood-based reasoning in infants.

8 Discussion

Popper's views on scientific inference seems derived from and more suitable for the hard sciences, particularly physics. Indeed, in his *Logic of Scientific Discovery,* he had a whole chapter on the quantum theory, while his propensity theory of probability applies naturally to physical phenomena. He somewhat downgraded the role of experience, empiricism and induction in the scientific discovery process; for him, all observations are 'theory laden,' i.e. the theory comes before observations. Supposedly, without any theory to begin with, how would anyone even know what observations to collect? According to a perhaps apocryphal story, at the start of his course on philosophy of science he liked to tell his students, 'Go ahead and observe!' Then he would just sit and wait; this was meant to show that without any theory there was nothing to observe. Ironically, one thing all scientists must know is that it is dangerous to generalize from single episodes: how much can we say about the logic of scientific discovery from an observationally barren class-room?

It is instructive to contrast physics with truly empirical sciences such as economics or medicine, where observations are collected all the time, with a purpose for sure, but mostly without any theory. For instance, consider the cancer registration, which, in most countries, is mandated by law. Observations such as cancer incidence are collected without any theory in mind; they are simply done for the purpose of monitoring of disease burden and health planning. But a doctor reading through the cancer records may notice, that certain professions, for instance, chimney sweeps, have much higher rates of scrotal cancer, while miners have higher rates of lung cancer. One may conjecture theories what those risks are, but the original observations that lead to the theories were not themselves ´theory laden'. Endless hypotheses can be formed by going through cancer registry data that cannot be conjectured by pure thinking. In recent medical genetics research, the most successful approach is the hypothesis-free genome-wide studies. This era came after the much lamented fruitless decades of the so-called candidate-gene approach with 'theory-laden' observations. Popper's emphasis on the hard-science-based scientific discovery had created an unnecessarily hard demarcation between deductive and inductive logic. In this paper we describe statistical ideas that are to a large extent Popperian, but also contains the logical elements of inductivism as captured by the Fisherian and Bayesian statistics. Specifically, these are represented by the likelihood and confidence concepts.

With a regular use of the inverse probability method of Laplace, nineteenth-century statistics was largely Bayesian. Fisher (1930) criticized the use of inverse probability method due to its arbitrary presumption of prior probability, but his own solution preserved the logico-inductive content of probability. There is a virtual consensus regarding the use of probability for statistical modeling, but we have yet to reach that for its interpretation and philosophical aspects. Mathematically, Kolmogorov's axiomatic foundation puts probability as the legitimate child of the mature measure theory. Kolmogorov's probability is naturally interpreted as long-run frequency. Thus, in most statistical frequentist textbooks, probability is said to be meaningless for specific single events such as Donald Trump's impeachment or re-election. But people do bet on such specific events, which can only mean that they do have a logical probability that is not a long-term variety. The reasoning requires a logical probability that applies to specific single events. Moreover, since different people have different beliefs and temperaments, they may have different subjective logical probabilities for the same event even though they share the common information (evidence).

We have described alternative measures of uncertainty including classical likelihood and confidence, and highlighted the differences with Bayesian logical probability. To a large extent this is done in the Popperian spirit of non-probabilistic corroboration. But with confidence and extended likelihood, we are preserving the logico-inductive spirit of Fisher. Likelihood and probability are bread-and-butter concepts in routine statistical analyses of scientific data. What do we gain by distinguishing the (logical) probability from the (inductive) likelihood-based reasoning? Primarily, clarifying the meaning of the terminologies will also clarify our thinking, thereby reducing unnecessary confusions. For example, we believe there is no need to call the conjunction fallacy a fallacy, and to accuse ourselves as being illogical or irrational, when we are in fact using likelihood-based reasoning. Closing the gap between the technical meaning and layman understanding is always a difficult challenge in the public dissemination of science, but perhaps not hopeless. At the very least, the distinction between probability vs likelihood-based reasoning should be part of a standard scientific discourse on human decision making. The likelihood-based reasoning should be recognized as an objective and rational mode of reasoning. Recently, likelihood has been extended to allow the sense of uncertainty associated with a realized but still unobserved single event, while at the same time avoid potential probability-related paradoxes (Pawitan and Lee 2017). Extended likelihood and therefore the confidence is not necessarily probability, so that care is necessary when expected utility is computed using confidence (Pawitan and Lee 2020b).

Even in statistics after 100 years since its introduction, there is still no general consensus on the direct use of likelihood for inference, indicating that it is difficult to give a normative answer. In statistics literature we can point to Edwards (1992) and Royall (1997) as proponents of this mode of likelihood inference, although we must add that they do not represent the Fisherian views we state above. Fisher (1973) recognized the two logical levels of uncertainty, whereas in this article we explain four levels, namely, classical likelihood, confidence, logical probability and Kolmogorov mathematical probability. Extended likelihood is proposed for simultaneous inferences of fixed and random unknowns (Lee et al. 2017). Classical likelihood is that for fixed unknowns, whereas confidence is that for binary random unknowns. We hope that these distinction enrich methodological developments in many non-statistical areas.

References

Bayes, T.: An essay towards solving a problem in the doctrine of chances. Phil. Trans. R. Soc. **53**, 370–418 (1763)

Bernardo, J.M.: Reference posterior distributions for Bayesian inference (with discussion). J. Roy. Stat. Soc.: Ser. B (Methodol.) **41**(2), 113–147 (1979)

Broad, C. D.: On the relation between induction and probability (Part I.). Mind **27**, 389–404 (1918)

Broad, C.D.: Scientific thought. Brace and Co, New York (1923)

Carnap, R.: Logical foundation of probability. University of Chicago Press (1950)

Cox, R.T.: Probability, frequency, and reasonable expectation. Am. J. Phys. **14**, 1–13 (1946)

Crupi, V., Fitelson, B., Tentori, K.: Probability, confirmation, and the conjunction fallacy. Think. Reason. **14**, 182–199 (2008)

De Finetti, B.: Theory of probability, vol. 1. John Wiley & Sons, London (1974)

Deutsch, D.: The beginning of infinity. Penguin Books (2011)

Edwards, A.W.F.: Likelihood, expanded edn. Johns Hopkins University Press, Baltimore (1992)

Etz, A., Wagenmakers, E.J.: JBS Haldane's contribution to the Bayes factor hypothesis test. Stat. Sci. **32**(2), 313–329 (2017)

Fisher, R.A.: On the "probable error" of a coefficient of correlation deduced from a small sample. Metron. **1**, 3–32 (1921)

Fisher, R.A.: Inverse probability. Proc. Camb. Philos. Society. **26**, 528–535 (1930)

Fisher, R.A.: The nature of probability. Centen. Rev. **2**, 261–274 (1958)

Fisher, R.A.: Statistical methods and scientific inference, 3rd edn. Hafner, New York (1973)

Gardner-Medwin, T.: What probability should a jury address? Significance **2**, 9–12 (2005)

Gillies, D.: Philosophical theories of probability. Routledge (2000)

Gödel, K.: Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. Monatshefte für mathematik und physik **38**(1), 173–198 (1931)

Gweon, H., Tenenbaum, J.B., Schulz, L.E.: Infants consider both the sample and the sampling process in inductive generalization. Proc. Natl. Acad. Sci. **107**(20), 9066–9071 (2010)

Hume, D.: An enquiry concerning human understanding. P.F. Collier & Son, (1910, original work published 1748)

Jaynes, E.T.: Probability theory: the logic of science. Cambridge university press (2003)

Jeffreys, H.: Theory of probability. Oxford University Press (1939)

Kant, I.: Critique of pure reason. Trans. Pluhar, W. Hackett, Indianapolis (1996, original work published 1781)

Kahneman, D.: Thinking, fast and slow. Penguin books (2011)

Keynes, J.M.: A treaties on probability. Macmillan, London (1921)

Kolmogorov, A.N.: Foundations of the theory of probability. 2nd English ed. Trans. Morrison, N. Chelsea Publishing Company, New York (1956, original work published 1933)

Laplace, P.S.: A philosophical essay on probabilities, 6th ed. Trans. Truscott, F. W., Emory, F. L. Dover, New York (1951, original work published 1814)

Lee, Y.: Resolving the induction problem: can we state with complete confidence via induction that the sun rises forever? arXiv:2001.04110 (2020)

Lee, Y., Nelder, J.A.: Hierarchical generalized linear models (with discussion). J. Roy. Stat. Soc.: Ser. B (Methodol.) **58**, 619–678 (1996)

Lee, Y., Nelder, J. A., Pawitan, Y.: Generalized linear models with random effects: unified analysis via H-likelihood, 2nd ed. Chapman & Hall (2017)

Neyman, J., Pearson, E.S.: On the problem of the most efficient tests of statistical hypotheses. Philos. Trans. R. Soc. Lond.: Ser. (Math. Phys. Character). **231**, 289–337 (1933)

Paulos, J.A.: The mathematics of changing your mind. New York Times, 5 August 2011

Pawitan, Y.: In all likelihood: statistical modelling and inference using likelihood. Oxford University Press (2001)

Pawitan, Y., Lee, Y.: Wallet game: probability, likelihood, and extended likelihood. Am. Statistician. **71**, 120–122 (2017)

Pawitan, Y., Lee, Y.: Confidence as likelihood. statistical science (2020a, In press)
Pawitan, Y., Lee, Y.: On rereading savage. A manuscript prepared for publication (2020b)
Popper, K.: New appendices. In: Popper, K. (ed.) The logic of scientific discovery. 6th revised impression of the 1959 English translation. Hutchinson (1972, original work published 1935)
Ramsey, F.: Truth and probability. In: Ramsey, F., Braithwaite, R.B. (eds.) Foundations of mathematics and other logical essays. Harcourt, New York (1931)
Royall, R.M.: Statistical evidence. Chapman and Hall, London (1997)
Russell, B.: The problems of philosophy. Henry Holt and Company (1912)
Savage, L.J.: Foundations of statistics. John Wiley & Sons (1954)
Schweder, T., Hjort, N.L.: Confidence, likelihood, probability. Cambridge University Press (2016)
Senn, S.: Comment on "Harold Jeffreys's theory of probability revisited". Statistical Science (2009)
Turing, A.M.: On computable numbers, with an application to the Entscheidungsproblem. Proc. Lond. Math. Soc. **42**, 230–265 (1936)
Tversky, A., Kahneman, D.: Extension versus intuitive reasoning: the conjunction fallacy in probability judgment. Psychol. Rev. **90**, 293–315 (1983)
Von Mises, R.: Probability, statistics and truth, 2nd revised English ed. Allen & Urwin (1961, original work published 1928)
Wittgenstein, L.: Logisch-Philosophische Abhandlung (Tractatus Logico-Philosophicus). Trinity College, Cambridge (1921)
Xie, M.G., Singh, K.: Confidence distribution, the frequentist distribution estimator of a parameter: a review. Int. Stat. Rev. **81**, 3–39 (2013)

Youngjo Lee is a Professor in the Department of Statistics at Seoul National University, South Korea. With the late Professor John Nelder, he introduced the hierarchical generalized linear models with unobserved random variables and proposed the use of an extended likelihood for inferences of random unknowns. His current research interests are extension, application and software development for extended likelihood approach. Recently, he devoted to writing expository papers on the meaning and contribution of extended likelihood in induction with the likelihood theorist Yudi Pawitan. E-mail: youngjo@snu.ac.kr

Yudi Pawitan is a Professor of biostatistics at Karolinska Institutet in Stockholm, Sweden since 2001. Having received his BSc in statistics at Bogor Agricultural Institute in Indonesia in 1982, he continued for a PhD in mathematical statistics at the University of California, Davis, and graduated in 1987. He has worked in many areas of statistical applications, including time series analyses and medical imaging, and for the last 20 years in the modelling and analysis of high-throughput molecular data arising from measurements of DNA, RNA and proteins. Statistical inference and philosophy of science have been a lifelong interest. For further information, please go to http://www.meb.ki.se/~yudpaw

Popper on Quantification and Identity

David Binder and Thomas Piecha

1 Introduction

Karl Popper developed a new approach to mathematical logic with foundational aspirations in the 1940s, which was published in a series of articles between 1946 and 1949. This new system of logic did not have the influence that he had hoped for, despite being original, and despite anticipating problems which were discussed in the logic community only much later. In a previous article (Binder and Piecha 2017) we explored in technical detail his approach to propositional logic, modal logic and various sub-classical systems like intuitionistic, dual-intuitionistic and minimal logic. A detailed discussion of his theory of quantification (i.e., of first-order logic) has, with the exception of an appendix to an article by Schroeder-Heister (1984), been lacking so far. We first present the main ideas of Popper's approach and the core of the propositional system. We then provide a concise introduction to his theory of quantification and identity, accessible to non-specialists. Popper's theory of quantification underwent significant modifications over the course of his published articles, subsequent corrections to those articles, and in unpublished correspondence with other logicians. We present what we consider to be his most mature view on these matters, taking unpublished material into account.

Popper's approach to logic is original, philosophically interesting, and also severely underappreciated. There are only a few detailed expositions and discussions of Popper's works on logic (cf. Schroeder-Heister 1984, 2006; Binder and Piecha 2017). Moreover, Popper's ideas on quantification have not yet received an extensive discussion, and in this article we would like to provide one. We first give a brief sketch of the genesis of Popper's ideas on logic in Sect. 2. In Sect. 3 we introduce the central philosophical ideas of Popper's approach to logic, namely to define logical constants by inferential definitions that are based on a deducibility relation.

D. Binder · T. Piecha (✉)
Department of Computer Science, University of Tübingen, Sand 13,
72076 Tübingen, Germany
e-mail: piecha@informatik.uni-tuebingen.de

D. Binder
e-mail: david.binder@uni-tuebingen.de

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_8

These ideas are exemplified by inferential definitions of connectives of propositional logic. In Sect. 4 we show how Popper intended to extend the propositional system to first-order logic (Popper uses the terms “theory of quantification” or “quantification theory” instead of “first-order logic”). At first, he extends his concept of object language to include open statements and his deducibility relation to range over open statements. He then adds a substitution operation which replaces free variables by other free variables, and gives rules and postulates which characterize this substitution operation. We discuss his definitions of the auxiliary concepts of identity and non-free-occurrence of a variable in a statement and, finally, his definitions of the quantifiers. We conclude in Sect. 5.

2 The Genesis of Popper’s Ideas on Logic

In January 1937 Karl Popper arrived in New Zealand and settled down in Christchurch, where he had found employment as a lecturer of philosophy at Canterbury University College. It is in Christchurch where he worked on, and finished, what he considered to be his contribution to the war, “The Open Society and Its Enemies” (Popper 1945). Combining teaching and research proved to be very difficult, and in his autobiography (Popper 1974) he complains about how the leadership of the university actively discouraged research which was not directly related to his teaching activities. But, as he also writes in his autobiography, he found the time to work on logic (Popper 1974, Sect. 25). While he only started to publish his work once he had returned to Europe and worked at the London School of Economics, it is clear that most of the genesis of his novel ideas on logic can be traced back to his time in Christchurch.

Since the university library in Christchurch was poorly equipped, Popper also relied on the personal library of Henry George Forder, who taught mathematics at Auckland University College and who lent him journal articles and monographs that Popper needed for his logical and mathematical research. Popper started an extensive correspondence with Forder in 1943 which mostly turned around questions of the foundations of physics, mathematics and logic. The correspondence with his pre-war contacts from Europe, on the other hand, proved to be difficult and slow. He did keep in contact with Carnap, who taught at the University of Chicago and who sent Popper his latest publications in logic. Popper writes to Carnap to tell him that he received the “Introduction to Semantics” (Carnap 1942) in October 1942 and “The Formalization of Logic” (Carnap 1943) at the end of June or beginning of July 1943.

A significant part of his time was spent on preparing the courses that he taught, one of them being the introduction to formal logic. Popper was always keen on expressing his opinions as clearly as he could, and this attitude also applied to formal logic. We think that it is likely that the teaching of logic to his Christchurch students was the occasion which prompted Popper to write down his thoughts on the foundations of logic. This is evidenced by the fact that he explicitly mentions discussions that he had with his student Peter Munz during one of his logic lectures in Christchurch (Popper 1974, Sect. 27 and endnote 194). After moving to England and taking up his new position at the London School of Economics he published the results obtained in New Zealand in a series of articles (Popper 1947c, a, d, 1948b, c, 1949). At the same time he also thought about writing a textbook on logic that he could use in his lectures. He writes about this plan in a draft of a letter to Alexander Carr-Saunders, the director of the London School of Economics at the time:

> I may say that I am at present preparing a textbook on formal logic, not because I like writing a textbook (it interferes, on the contrary, badly with my own research programme) but because I find it necessary for my students. The existing textbooks have aims totally different from what I consider to be the aim of a modern introductory course in Logic (Popper 1946).

Indeed, already in 1939/41 Popper had prepared lecture notes on logic (Popper 1941), and a table of contents for a textbook on logic can be found in Popper's estate (Popper n.d.b). Moreover, together with Paul Bernays he wrote a manuscript "On Systems of Rules of Inference" (Popper and Bernays n.d) which contains an exposition of Popper's original approach to logic. The jointly written manuscript was not published, however.

3 Inferential Definitions

In his approach to logic, Popper considers pairs of an *object language* $\mathcal{L}$ and a *deducibility relation* (also called *derivability relation*), written $/$, defined on $\mathcal{L}$. A given object language need not be a formal language but can also be a natural language. The deducibility relation between statements $a_1, \ldots, a_n$ and b is written as

$$a_1, \ldots, a_n / b$$

and is characterized by a so-called *basis*. Popper uses different bases. For clarity, we will use the following simple basis from Popper (1948b):

$$a_1, \ldots, a_n / a_i \quad (1 \leq i \leq n) \tag{Refl}$$

$$a_1, \ldots, a_n / b \rightarrow (b, a_1, \ldots, a_n / c \rightarrow a_1, \ldots, a_n / c) \tag{Trans}$$

The basis is formulated in a symbolic metalanguage, where → stands for "if-then". Further metalinguistic symbols are used, with the following meanings:

Symbol	$\rightarrow$	$\leftrightarrow$	&	(a)
Meaning	if-then	if and only if	and	for all a

Note that the axioms (Refl) and (Trans) are thus metalinguistic statements about the deducibility relation. They express that the deducibility relation $/$ is reflexive and transitive. Besides these two structural properties nothing else characterizes the primitive notion of deducibility.

Popper distinguishes between a *general theory of derivation*, which deals with deducibility and related notions, and a *special theory of derivation*, in which logical constants are defined in terms of deducibility.

For example, in the general theory the relation of *mutual deducibility* // is defined in terms of deducibility / as follows:

$$a//b \leftrightarrow (a/b \,\&\, b/a) \tag{mutual deducibility}$$

This is an equivalence relation, and two mutually deducible statements a and b are said to have the same *logical force*. Thus, the equivalence classes induced by // are logical forces. Another important defined relation is *relative demonstrability*, written $a_1, \ldots, a_n \vdash b_1, \ldots, b_m$:

$$a_1, \ldots, a_n \vdash b_1, \ldots, b_m \leftrightarrow (c)((b_1/c \,\&\, \ldots \,\&\, b_m/c) \rightarrow a_1, \ldots, a_n/c) \tag{relative demonstrability}$$

In words: The statements $b_1, \ldots, b_m$ are demonstrable relative to statements $a_1, \ldots, a_n$ (by definition) if, and only if, for all statements c: if c is deducible from each of the statements $b_1, \ldots, b_m$, then c is deducible from the statements $a_1, \ldots, a_n$ taken together. The notion of relative demonstrability is especially useful in cases where the object language contains conjunction $\wedge$ and disjunction $\vee$, since one can then show

$$a_1, \ldots, a_n \vdash b_1, \ldots, b_m \leftrightarrow a_1 \wedge \ldots \wedge a_n \vdash b_1 \vee \ldots \vee b_m$$

which gives us an interpretation of Gentzen's sequents (cf. Gentzen 1935a). From this point of view, Popper's basis characterizes commas on the left side of $\vdash$ as conjunction and commas on the right side of $\vdash$ as disjunction. Furthermore, the notion of relative demonstrability contains as special cases the concepts of complementarity, demonstrability, contradictoriness and refutability, which Popper defines as well (cf. Binder and Piecha 2017 for details).

The primitive notion of deducibility (and the notions defined in terms of it) is the foundation of Popper's *special theory of derivation*. In this theory, logical constants are defined in terms of deducibility alone. That is, a sign of a given object language is a logical constant, if, and only if, the sign can be defined by deducibility. Such definitions of logical constants (or *formative signs*, as Popper also calls them) are called *inferential definitions* by Popper:

> [...] inferential definitions [...] are characterized by the fact that they define a formative sign by its logical force which is defined, in turn, by a definition in terms of inference (i.e., of "/") (Popper 1947a, p. 286).

Inferential definitions of logical constants have the following form (where we use $\circ$ as a placeholder for an arbitrary binary connective):

$$a//a_1 \circ a_2 \leftrightarrow \mathcal{R}(a, a_1, a_2) \tag{D $\circ$}$$

In words: The object language statement a has the same logical force as the complex object language statement $a_1 \circ a_2$ if, and only if, the condition $\mathcal{R}(a, a_1, a_2)$ holds. Condition $\mathcal{R}(a, a_1, a_2)$ is a formula of the (symbolic) metalanguage containing (among others) the statements a, a_1, a_2 and the deducibility relation / (or maybe relations like $\vdash$, which are defined in terms of /). Popper calls a definition of the form (D$\circ$) an *explicit definition* of the connective $\circ$. To simplify the presentation one can consider only the right part of such definitions, replacing a by $a_1 \circ a_2$ in $\mathcal{R}$:

$$\mathcal{R}(a_1 \circ a_2, a_1, a_2) \tag{C $\circ$}$$

This is called the *characterizing rule* (C$\circ$); it corresponds to the definition (D$\circ$).

As examples, we show some inferential definitions of connectives given by Popper:

Conjunction $\wedge$:

$$a//b \wedge c \leftrightarrow (d)(a \vdash d \leftrightarrow b, c \vdash d) \tag{D $\wedge$}$$

$$b \wedge c \vdash d \leftrightarrow b, c \vdash d \tag{C $\wedge$}$$

Disjunction $\vee$:

$$a//b \vee c \leftrightarrow (d)(d \vdash a \leftrightarrow d \vdash b, c) \tag{D $\vee$}$$

$$d \vdash b \vee c \leftrightarrow d \vdash b, c \tag{C $\vee$}$$

Conditional $>$:

$$a//b > c \leftrightarrow (d)(d \vdash a \leftrightarrow d, b \vdash c) \tag{D $>$}$$

$$d \vdash b > c \leftrightarrow d, b \vdash c \tag{C $>$}$$

Popper also considers several definitions for *classical negation* ($\neg_k$), among them the following two, which are equivalent:

$$a// \neg_k b \leftrightarrow (a, b \vdash \ \& \ \vdash a, b) \tag{D $\neg_k$ 1}$$

$$a// \neg_k b \leftrightarrow (c)(d)(d, a \vdash c \leftrightarrow d \vdash b, c) \tag{D $\neg_k$ 2}$$

The characterizing rules are the following:

$$\neg_k b, b \vdash \ \& \ \vdash \neg_k b, b \tag{C $\neg_k$ 1}$$

$$(c)(d)(d, \neg_k b \vdash c \leftrightarrow d \vdash b, c) \tag{C $\neg_k$ 2}$$

Other examples of unary connectives are the following:

Tautology t:

$$a//t(b) \leftrightarrow (c)(b/a \leftrightarrow c/a) \tag{D t}$$

$$(c)(b/t(b) \leftrightarrow c/t(b)) \tag{C t}$$

Contradiction f:

$$a//f(b) \leftrightarrow (c)(a/b \leftrightarrow a/c) \tag{D f}$$

$$(c)(f(b)/b \leftrightarrow f(b)/c) \tag{C f}$$

We have for all statements b: $\vdash t(b)$ and $f(b) \vdash$. In other words, t is a unary verum, and f is a unary falsum.

Popper's approach is not restricted to classical logic. For example, he inferentially defines several kinds of non-classical negations, such as

Intuitionistic negation $\neg_i$:

$$a /\!/ \neg_i b \ \leftrightarrow\ (c)(c \vdash a \ \leftrightarrow\ c, b \vdash) \tag{D $\neg_i$}$$

$$c \vdash \neg_i b \ \leftrightarrow\ c, b \vdash \tag{C $\neg_i$}$$

For Popper, the availability of a characterizing rule like $\mathcal{R}(c, a_1, \ldots, a_n)$ may not be a sufficient criterion for the logicality of the constant characterized by it. Thus an inferential definition of this form need not define a logical constant in all cases. As a stronger criterion for logicality, Popper considers the existence of so-called *fully characterizing rules*, which are characterizing rules satisfying uniqueness in the sense that one can show that any two statements satisfying such a rule are mutually deducible (i.e., have the same logical force). In other words, a rule $\mathcal{R}(c, a_1, \ldots, a_n)$ is called *fully characterizing* if, and only if,

$$\mathcal{R}(a, a_1, \ldots, a_n) \ \&\ \mathcal{R}(b, a_1, \ldots, a_n) \to a /\!/ b.$$

The existence of fully characterizing rules is then used to distinguish between logical and non-logical constants (cf. the discussion in Schroeder-Heister (1984, 2006) and Binder and Piecha (2017, Sect. 4.3).

4 Substitution, Identity and Quantification

We cannot say precisely when Popper's ideas about propositional logic took shape. In the introduction to "New Foundations for Logic" (Popper 1947d) he writes that he obtained the results "during the last ten years", that is, between 1937 and 1947, roughly corresponding to the time he spent in New Zealand. On the other hand, we can give the exact date when he extended his inferential definitions to quantifiers. In a letter to Paul Bernays dated October 19th 1947 (Popper 1947f) he writes:

> The first important result which I had finished about one week after I saw you, was the extension of the method of $a/b \wedge c \ \leftrightarrow\ a/b \ \&\ a/c$ to quantification.

The meeting that Popper refers to probably took place in Zürich on April 11th or 12th 1947,[1] where Popper met Bernays in order to discuss the possibility of publishing a joint article on logic. The manuscript (Popper and Bernays n.d) for this unpublished article does not have a title; in a letter to Bernays dated March 3rd 1947 (Popper 1947e), Popper suggests the title "On Systems of Rules of Inference", noting that "[t]he title is not very good, but so far I could not think of a better one".

---

[1] Bernays writes to Popper: "[…] nothing stands, as far as I can see, in the way of us seeing each other on April 11th in Zurich; I will certainly also be available in the midmorning of the 12th. I'm looking forward to the receipt of the concept you promised me,—also with regards to the possible joint publication" (Bernays 1947).

Although they did not publish this manuscript, Popper's results found their way into several of his published articles. The most extensive discussion of these results can be found in Sect. 7 and 8 of "New Foundations for Logic" (Popper 1947d). Additionally, there is an important footnote in Popper (1948c), an alternative axiomatization in Popper (1947a), and a very short but clear summary of his treatment of quantification in Popper (1949). We follow the presentation of "New Foundations" (Popper 1947d) but refer to some modifications which can be found in his other articles. Some modifications of his view on quantification were only discussed in hitherto unpublished correspondence,[2] which we will discuss in Sect. 4.4.

4.1 Formulas, Name-Variables and Substitution

For propositional logic, as we saw in Sect. 3, Popper considered pairs

$$(\mathcal{L}; a_1, \ldots, a_n/b)$$

of an object language $\mathcal{L}$ and a deducibility relation $/$, axiomatized by a basis consisting of the rules (Refl) and (Trans). Each element of the object language $\mathcal{L}$ was presumed to be a statement, that is, something which can have a truth value.

The first modification Popper makes in order to treat quantification is to consider 4-tuples

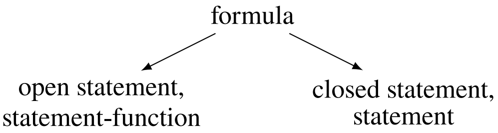
$$(\mathcal{L}; \mathcal{P}; a_1, \ldots, a_n/b; a\tbinom{x}{y})$$

consisting of a set $\mathcal{L}$ of *formulas*, a set $\mathcal{P}$ of *name-variables* (or pronouns), a *deducibility relation* on $\mathcal{L}$ and a *substitution operation*

$$a\tbinom{x}{y}$$

which substitutes the name-variable y for the name-variable x in the formula a. Variables $a, b, \ldots$ now range over *formulas* in $\mathcal{L}$, and variables $x, y, \ldots$ range over name-variables in $\mathcal{P}$.

Formulas can either be *open statements* (also called *statement-functions*) or *closed statements* (also called *statements*):



An example of an open statement given by Popper is "He is a charming fellow", which can be turned into a closed statement by replacing the name-variable "He" by the name "Ernest's best friend". Popper explicitly remarks that open statements do not have a truth value on their own; an open statement cannot be considered to be true or false.

The deducibility relation is axiomatized by the same rules (Refl) and (Trans) as in the case of propositional logic, but it now ranges over arbitrary formulas, not just closed statements.

---

[2] To be published in Binder et al. (2021b).

For example, Popper says that one can validly deduce the open statement “He is an excellent physician” from the open statement “He is not only a charming fellow but an excellent physician”.

The new substitution operation is characterized by the four postulates (PF1) to (PF4) and the six primitive rules of derivation (6.1) to (6.6), which we present in a slightly simplified form in the following.

$$\mathcal{L} \cap \mathcal{P} = \emptyset \tag{PF1}$$

$$\text{If } a \in \mathcal{L} \text{ and } x, y \in \mathcal{P}, \text{ then } a\tbinom{x}{y} \in \mathcal{L} \tag{PF2}$$

$$\text{For all } a \in \mathcal{L} \text{ there exists an } x \ \in \mathcal{P} \text{ such that for all } y \in \mathcal{P} : a\tbinom{x}{y}//a \tag{PF3}$$

$$\text{There exist } a \ \in \mathcal{L} \text{ and } x, y \in \mathcal{P} : a/a\tbinom{x}{y} \rightarrow t/f \tag{PF4}$$

Note that two kinds of metalinguistic quantifiers are used: There are universal and existential quantifiers ranging over statements $a \in \mathcal{L}$ and universal and existential quantifiers ranging over name-variables $x \in \mathcal{P}$. We only use symbols for the respective metalinguistic universal quantifiers in the following; (a) means “for all statements a” and (x) means “for all name-variables x”.

The postulates (PF1) and (PF2) are, in a way, only about the correct grammatical use of formulas and name-variables. The postulate (PF3) says that for every formula there is some name-variable not occurring in it. This is obvious if the set of name-variables is considered to be infinite, and if each formula is a finite object which can only mention a finite number of name-variables. The postulate (PF4), which Popper considers to be optional, excludes degenerate systems in which only one object exists. Take, for example, the open statement a to be “x likes the current weather”. The deducibility of “y likes the current weather” from “x likes the current weather” only leads to a contradiction if there are at least two persons to whom x and y can refer. Postulate (PF4) was also discussed in correspondence between Popper and Carnap (Carnap 1947; Popper 1947b; cf. Appendix A and B).

The six primitive rules of inference are given below. We will not discuss them in detail, but the reader may check that they are valid for a concrete formalized object language and a substitution operation for that language.

$$\text{If, for every} z,\ a//a\tbinom{y}{z} \text{ and } b//b\tbinom{y}{z}, \text{ then } a//b \rightarrow a\tbinom{x}{y}//b\tbinom{x}{y} \tag{6.1}$$

$$a//a\tbinom{x}{x} \tag{6.2}$$

$$\text{If } x \neq y, \text{ then } (a\tbinom{x}{y})\tbinom{x}{z}//a\tbinom{x}{y} \tag{6.3}$$

$$(a\tbinom{x}{y})\tbinom{y}{z}//(a\tbinom{x}{z})\tbinom{y}{z} \tag{6.4}$$

$$(a\binom{x}{y})\binom{z}{y}/\!/(a\binom{z}{y})\binom{x}{y} \tag{6.5}$$

$$\text{If } w \neq x, x \neq u \text{ and } u \neq y, \text{ then } (a\binom{x}{y})\binom{u}{w}/\!/(a\binom{u}{w})\binom{x}{y} \tag{6.6}$$

The rules (6.1) to (6.6) characterize substitution as a structural operation; this is similar to how the basis characterizes commas in sequences of statements. It is remarkable that Popper here presents an algebraic treatment of substitution, which can be compared to the theory of explicit substitution developed much later (cf., e.g., Abadi et al. 1991).

As an intriguing sidenote, Popper compares the definition of substitution by the rules (6.1) to (6.6) to the definition of conjunction via the inferential definition (D$\wedge$). He writes:

> These six primitive rules determine the meaning of the symbol "$a\binom{x}{y}$" in a way precisely analogous to the way in which, say, [rule (D$\wedge$) determines] the meaning of conjunction [. . .] with the help of the concept of derivability "/" (Popper 1947d, p. 226).

However, it has been pointed out by Schroeder-Heister (1984, p. 106) that Popper's rules for substitution "cannot be brought into the form of an inferential definition of an operator of the object language". Hence, substitution cannot have the status of a logical constant according to Popper's criterion for logicality; his rules for substitution do not have the form of characterizing rules (and, consequently, no fully characterizing rules can be given either). Indeed, Popper also explains substitution as follows:

> The notation
>
> $$\text{“}a\binom{x}{y}\text{”}$$
>
> will be used as a (variable) metalinguistic name of the statement which is the result of substituting, in the statement a (open or closed), the variable y for the variable x, wherever it occurs. $a\binom{x}{y}$ is identical with a if x does not occur in a Popper (1947a, p. 1216).

Popper's rules for substitution may thus be viewed as "an implicit characterization of a metalinguistic operation" (Schroeder-Heister 1984, p. 106), and not as an inferential definition of a logical constant for object languages.

Next we discuss some auxiliary concepts defined with the help of both the deducibility relation and the substitution operation.

4.2 Non-dependence, Identity and Difference

If we work with some inductively defined formal object language, then we can easily specify the set of free variables of a formula by recursion on the structure of that formula. This possibility is excluded in Popper's approach, since its generality does not restrict us to the consideration of formal languages. Popper therefore defines the concept

$$a_{\grave{x}}$$

which can be read as "x does not occur among the free variables in a". Popper himself expresses this as "a does not depend on x", "*a-without-x*" and "x does not occur relevantly in a". The formula a does not depend on x if, and only if, the substitution of some name-variable y for x does not change the logical strength of a. That is:

$$a /\!/ a_{\grave{x}} \leftrightarrow \text{ for every } y : a /\!/ a\tbinom{x}{y} \tag{D $a_{\grave{x}}$}$$

The second concept Popper defines with the help of deducibility and substitution is *identity*. As Popper (1947d, p. 227f, fn 24) notes, one first has to extend the object language $\mathcal{L}$ to incorporate formulas of the form *Idt*(x, y); this is achieved by the postulate

$$\text{If } x \text{ and } y \text{ are name variables, then } \mathit{Idt}(x, y) \text{ is a formula} \tag{P Idt}$$

In addition, the characterizing rules for substitution have to be extended by rules of the form

$$(\mathit{Idt}(x, y))\tbinom{x}{z} /\!/ \mathit{Idt}(z, y) \tag{A}$$

$$(\mathit{Idt}(x, y))\tbinom{y}{z} /\!/ \mathit{Idt}(x, z) \tag{B}$$

$$\text{If } x \neq u \neq y, \text{ then } \mathit{Idt}(x, y)\tbinom{u}{z} /\!/ \mathit{Idt}(x, y) \tag{C}$$

With these preliminaries, Popper defines identity using the following idea:

> The identity statement “*Idt*(x, y)” can be defined as the weakest statement strong enough to satisfy the […] formula […]
>
> “*Idt*$(x, y), a(x)/a(y)$”
>
> that is to say, the formula corresponding to what Hilbert-Bernays call the second identity axiom. (Hilbert-Bernays’s first axiom follows from the demand that the identity statement must be the *weakest* statement satisfying this formula.) (Popper 1949, p. 725f)

The identity axioms Popper refers to are the axioms J_1 and J_2 of Hilbert (1934, p. 164):

$$a = a \tag{J$_1$}$$

$$a = b \rightarrow (A(a) \rightarrow A(b)) \tag{J$_2$}$$

This justifies the following definition of *identity* *Idt*(x, y):

$$\begin{aligned} a /\!/ \mathit{Idt}(x, y) \leftrightarrow {} & (\text{for every } b \text{ and } z\text{: } ((b /\!/ b_{\grave{x}} \;\&\; b /\!/ b_{\grave{y}}) \rightarrow a, b\tbinom{z}{x} / b\tbinom{z}{y}) \;\& \\ & ((\text{for every } c \text{ and } u\text{: } ((c /\!/ c_{\grave{x}} \;\&\; c /\!/ c_{\grave{y}}) \rightarrow b, c\tbinom{u}{x} / c\tbinom{u}{y}))) \rightarrow b/a)) \end{aligned} \tag{D Idt}$$

Popper (1948c, p. 323f, fn 11) expands on the definition of identity *Idt*(x, y) in order to illustrate his method of obtaining a relatively simple characterizing rule from an explicit definition that is the weakest (or strongest) statement satisfying a certain condition or axiom. He first introduces the following abbreviating notation:

$$a /\!/ a_{\grave{x}\,\grave{y}} \leftrightarrow (w)(a /\!/ a\tbinom{x}{w} \;\&\; a /\!/ a\tbinom{y}{w}).$$

Using this abbreviation, he defines *Idt*(x, y) as the weakest statement strong enough to imply the axiom J_2:

$$
\begin{aligned}
& a /\!/ \mathit{Idt}(x, y) \leftrightarrow \\
& \qquad (b)(z)((b /\!/ b_{\grave{x}\,\grave{y}} \rightarrow a, b\binom{z}{x}/b\binom{z}{y}) \mathbin{\&} (((c)(u)(c /\!/ c_{\grave{x}\,\grave{y}} \\
& \qquad \rightarrow b, c\binom{u}{x}/c\binom{u}{y})) \rightarrow b/a))
\end{aligned}
\tag{D $\mathit{Idt}^{\dagger}$}
$$

This explicit definition, which is an abbreviated version of (D *Idt*), can be replaced by a definition that corresponds to the following characterizing rule:

$$
a/\mathit{Idt}(x, y) \leftrightarrow (b)(z)(b /\!/ b_{\grave{x}\,\grave{y}} \rightarrow a, b\binom{z}{x}/b\binom{z}{y}) \tag{C $\mathit{Idt}^{\dagger}$}
$$

This can be seen by instantiating a in (D $\mathit{Idt}^{\dagger}$) with $\mathit{Idt}(x, y)$ in order to obtain

$$
\begin{aligned}
& (b)(z)((b /\!/ b_{\grave{x}\,\grave{y}} \rightarrow \mathit{Idt}(x, y), b\binom{z}{x}/b\binom{z}{y}) \mathbin{\&} \\
& \qquad\qquad (((c)(u)(c /\!/ c_{\grave{x}\,\grave{y}} \rightarrow b, c\binom{u}{x}/c\binom{u}{y})) \rightarrow b/\mathit{Idt}(x, y))).
\end{aligned}
$$

The left conjunct gives the direction from left to right in (C $\mathit{Idt}^{\ddagger}$), and the right conjunct gives the direction from right to left.

Finally, *difference* $\mathit{Dff}(x, y)$ is simply defined as the classical negation of identity:

$$
a /\!/ \mathit{Dff}(x, y) \leftrightarrow a /\!/ \neg_k \mathit{Idt}(x, y) \tag{D Dff}
$$

It is interesting to see that Popper chose to treat occurrence of free variables and identity as defined notions, rather than to class them with substitution and deducibility among the primitive notions characterized by the basis. We will see in cf. Sect. 4.4 that Popper probably revised this position later.

4.3 Quantification

Inferential definitions of universal and existential quantification are introduced in Popper (1947d), to which he later published a list of corrections and additions (Popper 1948a), which we take into account here. Popper's aim is not to develop and analyze the theory of quantification, that is, first-order logic, but to show that his approach to quantification is at least on a par with other proposed treatments of quantification. He therefore restricts himself to stating his definitions of the quantifiers and to deriving some simple conclusions, but he does not formally develop a meta-theory of quantification. He does not, for example, discuss the completeness of his rules, the difference between classical and constructive interpretations of the existential quantifier, or the relation to models of his system.

Later, Popper (1949) gives the clearest explanation of what intuition his inferential definition of *universal quantification* is supposed to capture. He writes:

> The result of universal quantification of a statement a can be defined as the weakest statement strong enough to satisfy the law of specification, that is to say, the law "what is valid for all instances is valid for every single one" (Popper 1949, p. 725).

Presupposing his rules of substitution, and writing Ax for the universal quantifier, Popper's inferential definition and the characterizing rule for universal quantification are the following:

$$a\grave{y}//Axb\grave{y} \leftrightarrow \text{(for every } c\grave{y} : c\grave{y}/a\grave{y} \leftrightarrow c\grave{y}/b\grave{y}\tbinom{x}{y}) \tag{D7.1}$$

$$\text{For every } c\grave{y} : c\grave{y}/Axb\grave{y} \leftrightarrow c\grave{y}/b\grave{y}\tbinom{x}{y} \tag{C7.1}$$

In order to see how more ordinary presentations of the rules for universal quantification follow from these inferential definitions, we can compare them to the more familiar rules of the (intuitionistic) sequent calculus (writing $\varphi[x/y]$ for the result of substituting y for x in the formula φ):

$$\frac{\Gamma, \varphi[x/t] \vdash \psi}{\Gamma, \forall x\varphi \vdash \psi}(\forall\vdash) \qquad \frac{\Gamma \vdash \varphi[x/y]}{\Gamma \vdash \forall x\varphi}(\vdash\forall)$$

$$\frac{\Gamma, \varphi[x/y] \vdash \psi}{\Gamma, \exists x\varphi \vdash \psi}(\exists\vdash) \qquad \frac{\Gamma \vdash \varphi[x/t]}{\Gamma \vdash \exists x\varphi}(\vdash\exists)$$

with the variable condition that y does not occur free in the conclusion of $(\vdash \forall)$ and $(\exists \vdash)$.

For example, by instantiating (C7.1) with $Axb\grave{y}$ and by using the rules (Trans) and (Refl) from the basis, we obtain the following rule

$$a, b\grave{y}\tbinom{x}{y}/c \rightarrow a, Axb\grave{y}/c$$

which can easily be seen to be a variant of the rule $(\forall \vdash)$ where the name-variable y takes the role of the term t. Similarly, by instantiating (C7.1) with $c\grave{y}$ and reading the biimplication from right to left we obtain the following rule, which corresponds to the rule $(\vdash \forall)$ with the variable condition that y does not occur relevantly in c:

$$c\grave{y}/b\grave{y}\tbinom{x}{y} \rightarrow c\grave{y}/Axb\grave{y}.$$

As was the case for universal quantification, Popper gives the clearest explanation of the inferential definition of *existential quantification* not in Popper (1947d), but in Popper (1949, p. 725):

> The result of existential quantification of the statement a can be defined as the strongest statement weak enough to follow from every instance of a.

The inferential definition and the characterizing rule for the *existential quantifier* Ex are

$$a\grave{y}//Exb\grave{y} \leftrightarrow \text{(for every } c\grave{y} : a\grave{y}/c\grave{y} \leftrightarrow b\grave{y}\tbinom{x}{y}/c\grave{y}) \tag{D7.2}$$

$$\text{For every } c\grave{y} : Exb\grave{y}/c\grave{y} \leftrightarrow b\grave{y}\tbinom{x}{y}/c\grave{y} \tag{C7.2}$$

To elucidate, we derive some more familiar rules for the existential quantifier from its characterizing rule. Instantiating (C7.2) with $Exb\grave{y}$ and using the rules of the basis we can obtain the rule

$$a/b\grave{y}\tbinom{x}{y} \rightarrow a/Exb\grave{y}$$

which corresponds to the sequent calculus rule ($\vdash \exists$); and by instantiating (C7.2) with $c\grave{y}$ and reading the biimplication from right to left, we obtain the following rule, which corresponds to ($\exists \vdash$):

$$b\grave{y}\binom{x}{y}/c\grave{y} \rightarrow Ex\grave{y}/c\grave{y}.$$

Popper does not consider the explicit definitions (D7.1) and (D7.2) to be improvements compared to the characterizing rules. They are given to show that universal and existential quantification can be defined using only his basis and the rules (6.1) to (6.6). He notices that these rules are not as simple as the rules of his basis, for example. But he points out that the concept of "$a\grave{x}$" can be avoided in these definitions (Popper 1947d, p. 230, fn 26, added in the corrections and additions Popper 1948a). Assuming $x \neq y$, one can use instead:

$$a\binom{y}{x}/Ax(b\binom{y}{x}) \leftrightarrow a\binom{y}{x}/b\binom{x}{y} \tag{7.1 *}$$

$$Ex(a\binom{y}{x})/b\binom{y}{x} \leftrightarrow a\binom{x}{y}/b\binom{y}{x} \tag{7.2 *}$$

$$a\binom{y}{x}/\!/Ax(b\binom{y}{x}) \leftrightarrow (\text{for every } c : c\binom{y}{x}/a\binom{y}{x} \leftrightarrow c\binom{y}{x}/b\binom{x}{y}) \tag{D7.1*}$$

$$a\binom{y}{x}/\!/Ex(b\binom{y}{x}) \leftrightarrow (\text{for every } c : a\binom{y}{x}/c\binom{y}{x} \leftrightarrow b\binom{x}{y}/c\binom{y}{x}) \tag{D7.2*}$$

He conceives his rules of quantification to be less complicated than those given by Hilbert and Ackermann (1928) or those given by Quine (1940, Sect. 15) , and he emphasizes that his rules in the end make use of only one logical concept, namely that of deducibility / as characterized by his basis.

4.4 An Unfortunate Misunderstanding

Popper (1947d, Sect. 8) introduces a distinction which he considered to be very important: the distinction between rules of derivation and rules of proof. If he had not stopped publishing in logic, then it is very likely that he would have developed these ideas in more detail. For example, among his unpublished manuscripts there are two which are entitled "Derivation and Demonstration in Propositional and Functional Logic" (Popper n.d.a) and "The Propositional and Functional Logic of Derivation and of Demonstration" (Popper n.d.c), as well as another untitled manuscript (Popper n.d.d), which also deals with the distinction between derivation and demonstration.

In order to illustrate this distinction we have to make use of the concept of relative demonstrability $\vdash$, which was introduced in Sect. 3. If we specialize this concept to no formula on the left hand side and exactly one formula on the right hand side, we obtain the definition of a *provable formula* a: $\vdash a$. Consider now the following two formulas of the metalanguage:

$$a/b \rightarrow (\vdash a \rightarrow \; \vdash b)$$
$$(\vdash a \rightarrow \; \vdash b) \rightarrow a/b.$$

Popper correctly remarks that while the first formula is valid, the second is not. This can be seen by instantiating a by a consistent formula and b by a contradictory one. Now Popper correctly observes that the rules of a system like Principia Mathematica (Whitehead and Russell 1927) are rules of proof and not rules of derivation. For example, the rule of modus ponens takes the form

$$\vdash a \rightarrow (\vdash a > b \rightarrow \ \vdash b)$$

rather than the form

$$a, a > b/b.$$

What Popper intends to formulate here, and in particular in his definition of a purely derivational system of primitive rules (cf. Popper 1947d, definition (D8.1)), is, in our opinion, a criterion that allows to distinguish between formulations of logic based on axioms and rules of proof, such as Hilbert and Bernays's 1934; 1939 or Whitehead and Russell's 1927 on the one hand, and formulations of logic based on derivation alone, such as Gentzen's 1935a; 1935b and his own, on the other hand.

Unfortunately, he applied this analysis of rules of derivation and rules of proof to the systems of Carnap as well as of Hilbert and Bernays in a way that does not take account of an important difference between his system and theirs. Popper (1947d, p. 232) warns that there are rules of proof such as

$$\vdash a\grave{y} \ \leftrightarrow \ \vdash a\grave{y}\binom{x}{y} \tag{8.5}$$

which are valid, whereas the corresponding rule of derivation

$$a\grave{y}/a\grave{y}\binom{x}{y}$$

is invalid. He continues:

> Now all the mistakes here warned against do actually vitiate some otherwise very excellent books on modern logic – an indication that the distinction between (conditional) rules of proof or rules of demonstration on the one side and rules of derivation on the other cannot be neglected without involving oneself in contradictions (Popper 1947d, p. 233).

Both Carnap and Bernays responded to Popper's criticism of their respective system in correspondence. We have reproduced Carnap's letter and Popper's response in Appendix A and B, respectively. Bernays (1948) writes:

> Now I have to comment upon your critique of the formulation of the all-schema, as it is given in the "Grundlagen der Math." Hilbert and Bernays (1934). I think of the passage p. 232–233 of your New Foundations. […] The contradiction that you derive, starting with the schema $a_{\grave{x}} > b/a_{\grave{x}} > Axb$ which you criticize, does not arise in the formalism of the "Grundl. der Math.", because the implication plays another role here than the "hypothetical" in your formalism.

We note that Popper's letter to Carnap (cf. Appendix B) is also interesting for the fact that it contains an expansion of his theory of quantification by presenting several logical laws of classical first-order logic.

Popper later revised his understanding of the interaction of substitution and deducibility. While his definitions are formulated using the weaker notion of interdeducibility, he then considered it necessary to use the stronger notion of identity of statements (Popper 1974, p. 171, endnote 198; reproduced in Appendix C).

Concerning possible future work on logic, Popper states in his reply (Popper 1948d) to Bernays's letter (Bernays 1948):

> I have also a number of new results – but I do not believe that I will ever dare again to publish something (except, maybe, an infinite sequence of corrections to my old publications)!

5 Conclusion

Popper's works on logic in the 1940s had no real influence on the further development of logic. This is despite the fact that he anticipated and had results on several issues in the area of philosophical logic which are still discussed today. We mention his inferentialist approach to logic, his analysis of logicality, and his results on combining logical systems (cf. Binder and Piecha 2017; Schroeder-Heister 1984, 2006 for details). In his inferentialist approach to logic, Popper anticipated many ideas of proof-theoretic semantics (Schroeder-Heister 2018; Piecha and Schroeder-Heister 2016; cf. Binder et al. 2021b).

At the time, his works were reviewed by several prominent logicians, including Ackermann (1948, 1949a, b), Beth (1948), Curry (1948a, b, c, d, 1949), Hasenjaeger (1949), Kleene (1948, 1949), McKinsey (1948). While the reviews by Ackermann and Beth are summarizing, Curry, Hasenjaeger, Kleene and McKinsey are critical about certain aspects of Popper's approach and point out some technical issues (for a discussion of these criticisms cf. Schroeder-Heister 1984). Concerning Popper's treatment of quantification in particular, Curry (1948a) raised some doubts (which were also discussed by Seldin 2008), although without going into details and while maintaining that "[p]resumably [Popper's] ideas can be carried through, at least in principle". Popper also disseminated his ideas in correspondence with Carnap, for example, who saw the importance of Popper's work on logic (Carnap 1947; cf. Appendix A). Brouwer (1947) reacted positively as well and presented Popper's articles (Popper 1947a, 1948b, c) to the Koninklijke Nederlandse Akademie van Wetenschappen for publication. However, although Popper's approach to logic is original and philosophically quite interesting, it did not receive the wider appreciation it deserves.

Acknowledgements Thomas Piecha was supported by the DFG project "Constructive Semantics and the Completeness Problem", DFG grant PI 1174/1-1. He would like to thank Zuzana Parusniková for her kind invitation and for organizing the Karl Popper Symposium at CLMPST 2019 in Prague, as well as David Miller and the Karl Popper Charitable Trust for their support. Both authors would like to thank Nicole Sager and Reinhard Lube from the Karl Popper-Sammlung in Klagenfurt for their help and Peter Schroeder-Heister for comments.

Appendix A Letter from Carnap to Popper, October 9th 1947 (Carnap 1947)

Santa Fe, N.M.,
P.O.B.
1214

Oct. 9, 1947

Dear Popper,

My best thanks for your letter of August & your kind judgement on my book.

I just read your "New Found." in "Mind". It is very interesting & contains a number of new & important results. It is an essential improvement in comparison with my previous attempts of defining the connection in terms of "consequence" (first in "Syntax" § 57, & later in "Formalization"). Among other things, your clear & simple analysis of the three kinds of negation is very valuable.

Your discussion on pp. 232f. is, unfortunately, so short that I am unable to understand it. I should like to understand it, especially since it is the basis of your objections against my rules. You say that the last formula on p. 232 leads to that on top of p. 233. How does it? Is the restriction $a_{\grave{x}}$ in the former but not in the latter no impediment? Further, you say that $a/\!/a\binom{x}{y}$ violates PF2. How does it? (I say that it violates your interpretation.) This is only a question, not an objection; I assume that your assertions concerning *your* system are correct.

However, I doubt very much whether your assertions (in the footnotes p. 232) of the non-validity of my rules 10 & 11 in D28-2 are correct. Note that rule 11, because of its restriction, does not lead to a proof of "$Px \supset (x)(Px)$" (which would indeed be wrong), although "$(x)(Px)$" is derivable from "Px", in distinction to your system. You have probably made the mistake of inadvertently transforming the interpretation & the rules of your system to mine (+ Hilbert's, etc.) Perhaps you have overlooked the following essential distinction. In my system (& Hilbert's, etc. but perhaps not in Princ. Math., & certainly not in your system), "Px" (as a separate formula) is interpreted (see, e.g., "Syntax", p. 22, par. 2) as meaning the same as "Py" & as "$(x)(Px)$". Therefore my rules are valid, if you doubt it, please give a counter-example, by using only *my* rules, not yours.

Feigl & Hempel (& his new wife) were here for a few weeks, & we had a very nice time together, with many interesting discussions, mostly on inductive logic.

We shall stay here until Xmas.

With best regards,
yours,
Rudolf Carnap

(Please, let's forget about titles.)

Appendix B Letter from Popper to Carnap, November 24th 1947 (Popper 1947b)

November 24th, 1947.

Dear Carnap,

I am sending you to-day an offprint of my "Logic without Assumptions", referred to in my "New Foundations" (note 1 on p. 203). Two more papers are on the way; I have been promised the offprints of one of them for next week, and I shall send you a copy at once.

I am overworked (8 h lecturing a week is too much if one does research – I wish I could get some time off for research, but I don't know how), and really quite exhausted.

You asked me in your last letter for a fuller explanation of my pp. 232f. (of my "New Foundations"). I suppose that it is the misprint on p. 233 ("PF2" should properly read "PF4") which created the difficulty, and that you will have found meanwhile what I meant. Still, here is a fuller explanation.

My contention is this.

Your statement (*Formalization*, p. 136) "that there exists a one-one correlation between the individuals and the natural numbers" indicates that it is your intention to construct a calculus which is *consistent with* my (much weaker) postulate PF4, i.e., with the demand that there exists *more than one individual.*

But with the assumption that there exists more than one individual, each of the following rules of your *Formalization* contradict:

C10 (i.e., D28-2, rule 10, on p. 137)
C11 (i.e., D28-2, rule 11, on p. 138)
C12 (i.e., D28-2, rule 12, on p. 138)
Cb (i.e., T28-4, case b, on p. 139).

For the proof of this contention, I shall make use of my own formalism. But the proof holds for your formalism as well; for your C-implication satisfies, on the basis of your *Introduction*, p. 64, P14-5; P14-8; and P14-11, all the rules which define my "… / …", i.e., the rules which I shall call "generalized reflexivity principle" and "generalized transitivity principle". To the latter, I shall refer as "Tg".

I shall also refer to the following principles ("a^k" is the classical negation of a):

(1.1) $a/b \to\ \vdash a > b$
(1.2) $a/b \to (\vdash a \to\ \vdash b)$
(1.3) $\vdash a > b \leftrightarrow\ \vdash b^k > a^k$
(1.4) $(a \wedge a^k) \vee b//b$
(1.5) $a//a^{kk}$
(1.6) $(Ax(a^k))^k//Exa$ (cp. your *d* and *e*, p. 139)

I begin with C10, which I write

(C10) $a/a\binom{x}{y}$, provided y is not bound in $a\binom{x}{y}$.

We obtain, always assuming that y is not bound in $a\binom{x}{y}$:

(C10.1) $\vdash a > a\binom{x}{y}$ (C10;1.1)
(C10.2) $\vdash (a\binom{x}{y})^k > a^k$ (C10;1;1.3)
(C10.3) $(a\binom{x}{y})^k/a^k$ (C10.2;1.1)
(C10.4) $\vdash (a\binom{x}{y})^k \to\ \vdash a^k$ (C10.;1.2)

Now we take "a" to be the name of an open statement (such as "$x + 1 = y$") which is *satisfiable but not* universally true. We obtain

(C10.5) $\vdash$ "$y + 1 \neq y$" $\to\ \vdash$ "$x + 1 \neq y$"(C10.4)

and, substituting further "$x + 1$" for "y" (we *may* confine this to the right hand side, but I shall do it throughout) we obtain

(C10.6) $\vdash$ "$(x + 1) + 1 \neq x + 1$" $\to\ \vdash$ "$x + 1 \neq x + 1$" (C10.5)

If there exists only *one* individual, then every statement of the form "… $\neq$ —" is false, and C10.6 is innocuous. But if there are more individuals than one, C10.6 gives rise to

(C10.7) "$x + 1 \neq x + 1$"

which is contradictory.

I now proceed to C11. This may be written:

(C11) $a \vee b / a \vee Axb$, provided x is not free in a.

We obtain, substituting “$Axc \wedge (Axc)^k$” for “a”:

(C11.1) $(Axc \wedge (Axc)^k) \vee b / (Axc \wedge (Axc)^k) \vee Axb$ (C11)
(C11.2) b / Axb (C11.1;1.4;Tg.)
(C11.3) $\vdash b > Axb$ (C11.2;1.1)
(C11.4) $\vdash (Axb)^k > b^k$ (C11.3;1.3)
(C11.5) $\vdash (Ax(a^k))^k > (a^k)^k$ (C11.4)
(C11.6) $(Ax(a^k))^k / a^{kk}$ (C11.5;1.1)
(C11.7) Exa / a (C11.6;1.5;1.6;Tg.)
(C11.8) Exa / Axa (C11.7;C11.2;Tg.)

But C11.8 is, clearly, satisfied only if there is not more than *one* individual.

I proceed to rule C12. This may be written

(C12) $a^k \vee b / (Exa)^k \vee b$, provided x is not free in b.

Substituting “$Axc \wedge (Axc)^k$” for “b” (as before under C11), we obtain:

(C12.1) $a^k / (Exa)^k$ (C12)
(C12.2) $b^{kk} / (Ex(b^k))^k$ (C12.1)
(C12.3) b / Axb (C12.2;1.5;1.6;Tg.)

But C12.3 is the same as C11.2, and has the same fatal consequences.

Rule Cb, of course, is also the same as C11.2 and C12.3.

The result of all this is:

(1) Rule Cb can be dropped altogether.
(2) The rules of derivation C10; C11; and C12 must be replaced by the corresponding conditional rules of proof, C′10’; C′11; C′12:

(C’10) $\vdash a \rightarrow\ \vdash a\binom{x}{y}$, provided y is not bound in $a\binom{x}{y}$.
(C’11) $\vdash a \vee b \rightarrow\ \vdash a \vee Axb$, provided x is not free in$a$.
(C’12) $\vdash a^k \vee b \rightarrow\ \vdash (Exa)^k \vee b$, provided x is not free in$b$.

The last two rules may be replaced by C″11 and C″12:

(C”11) $a\binom{x}{y}/b \rightarrow a\binom{x}{y}/Axb$, provided $x \neq y$
(C”12) $a/b\binom{x}{y} \rightarrow Exa/b\binom{x}{y}$, provided $x \neq y$.

These two rules, in turn, can be replaced by:

(C”’11) $a\binom{y}{x}/b\binom{x}{y} \rightarrow a\binom{y}{x}/Axb\binom{y}{x}$ $(x \neq y)$
(C”’12) $a\binom{x}{y}/b\binom{y}{x} \rightarrow Exa\binom{y}{x}/b\binom{y}{x}$ $(x \neq y)$

If we replace here “→” by “↔”, we obtain the rules which define the quantifiers, and from which, in the presence of the six rules defining “$a\binom{x}{y}$”, everything else can be obtained:

(C””) $a\binom{y}{x}/Axb\binom{y}{x} \leftrightarrow a\binom{y}{x}/b\binom{x}{y}$ $(x \neq y)$
(C””) $Exa\binom{y}{x}/b\binom{y}{x} \leftrightarrow a\binom{x}{y}/b\binom{y}{x}$ $(x \neq y)$

Appendix C Popper (1974, p. 171, endnote 198)

The mistake was connected with the rules of substitution or replacement of expressions: I had mistakenly thought that it was sufficient to formulate these rules in terms of interdeducibility, while in fact what was needed was identity (of expressions). To explain this remark: I postulated, for example, that if in a statement a, two (disjoint) subexpressions x and y are both, wherever they occur, replaced by an expression z, then the resulting expression (provided it is a statement) is interdeducible with the result of replacing first x wherever it occurs by y and then y wherever it occurs by z. What I should have postulated was that the first result is identical with the second result. I realized that this was stronger, but I mistakenly thought that the weaker rule would suffice. The interesting (and so far unpublished) conclusion to which I was led later by repairing this mistake was that there was an essential difference between propositional and functional logic: while propositional logic can be constructed as a theory of sets of statements, whose elements are partially ordered by the relation of deducibility, functional logic needs in addition a specifically morphological approach since it must refer to the subexpression of an expression, using a concept like identity (with respect to expressions). But no more is needed than the ideas of identity and subexpression; no further description especially of the shape of the expressions.

References

Abadi, M., Cardelli, L., Curien, P.L., Levy, J.J.: Explicit substitutions. J. Funct. Program. **1**, 375–416 (1991)

Ackermann, W.: Review of Functional Logic without Axioms or Primitive Rules of Inference. Zentralblatt für Mathematik und ihre Grenzgebiete **29**, 196 (1948)

Ackermann, W.: Review of On the Theory of Deduction, Part I. Zentralblatt für Mathematik und ihre Grenzgebiete **30**, 3 (1949a)

Ackermann, W.: Review of On the Theory of Deduction, Part II. Zentralblatt für Mathematik und ihre Grenzgebiete **30**, 101 (1949b)

Bernays, P.: Unpublished letter to K.R. Popper, March 12th 1947 (1947). Karl Popper-Sammlung, Universität Klagenfurt, Box 276, Folder 12. Translation by the authors

Bernays, P.: Unpublished letter to K.R. Popper, May 12th 1948 (1948). ETH Zürich, Paul Bernays Collection, Hs 975: 3653. Translation by the authors

Beth, E.W.: Review of new foundations for logic. Math. Rev. **9**, 130 (1948). http://www.ams.org/mathscinet-getitem?mr=0021924

Binder, D., Piecha, T.: Popper's notion of duality and his theory of negations. Hist. Philos. Log. **38**(2), 154–189 (2017)

Binder, D., Piecha, T., Schroeder-Heister, P.: Popper's Theory of Deductive Logic (2021a); Introduction to Popper's articles on logic in: Binder, Piecha, and Schroeder-Heister (2021b)

Binder, D., Piecha, T., Schroeder-Heister, P.: The Logical Writings of Karl Popper. Springer, Berlin (2021b)

Brouwer, L.E.J.: Letter to K.R. Popper. December 10th, 1947, handwritten (1947). Karl Popper-Sammlung, Universität Klagenfurt, Box 280, Folder 8

Carnap, R.: Introduction to Semantics. Harvard University Press, Cambridge (1942)

Carnap, R.: Formalization of Logic. Harvard University Press, Cambridge (1943)

Carnap, R.: Letter to K.R. Popper, October 9th 1947 (1947). Karl Popper-Sammlung, Universität Klagenfurt, Box 282, Folder 24

Curry, H.B.: Review of Functional Logic without Axioms or Primitive Rules of Inference. Math. Rev. **9**, 321 (1948a). http://www.ams.org/mathscinet-getitem?mr=0023207

Curry, H.B.: Review of Logic without Assumptions. Math. Rev. **9**, 486 (1948b). http://www.ams.org/mathscinet-getitem?mr=0024397

Curry, H.B.: Review of On the Theory of Deduction, Part I. Math. Rev. **9**, 486–487 (1948c). http://www.ams.org/mathscinet-getitem?mr=0024398

Curry, H.B.: Review of On the Theory of Deduction, Part II. Math. Rev. **9**, 487 (1948d). http://www.ams.org/mathscinet-getitem?mr=0024399

Curry, H.B.: Review of the Trivialization of Mathematical Logic. Math. Rev. **10**, 422 (1949). http://www.ams.org/mathscinet-getitem?mr=0028263

Gentzen, G.: Untersuchungen über das logische Schließen, I. Mathematische Zeitschrift **39**(1), 176–210 (1935a). https://doi.org/10.1007/BF01201353. English Translation in Gentzen (1969)

Gentzen, G.: Untersuchungen über das logische Schließen, II. Mathematische Zeitschrift **39**(1), 405–431 (1935b). https://doi.org/10.1007/BF01201363. English Translation in Gentzen (1969)

Gentzen, G.: In: Szabo, M.E. (ed.) The Collected Papers of Gerhard Gentzen. North-Holland, Amsterdam (1969)

Hasenjaeger, G.: Review of the Trivialization of Mathematical Logic. Zentralblatt für Mathematik und ihre Grenzgebiete **31**, 193 (1949)

Hilbert, D., Ackermann, W.: Grundzüge der theoretischen Logik. Springer, Berlin (1928)

Hilbert, D., Bernays, P.: Grundlagen der Mathematik, vol. 1. Springer, Berlin (1934)

Hilbert, D., Bernays, P.: Grundlagen der Mathematik, vol. 2. Springer, Berlin (1939)

Kleene, S.C.: Review of functional logic without axioms or primitive rules of inference. J. Symb. Log. **13**(3), 173–174 (1948)

Kleene, S.C.: Review of On the Theory of Deduction, Part I and II, and the Trivialization of Mathematical Logic. J. Symb. Log. **14**(1), 62–63 (1949)

McKinsey, J.C.C.: Review of Logic without Assumptions and New Foundations for Logic. J. Symb. Log. **13**(2), 114–115 (1948)

Piecha, T., Schroeder-Heister, P. (eds.): Advances in Proof-Theoretic Semantics. Trends in Logic, vol. 43. Springer, Berlin (2016). https://link.springer.com/content/pdf/10.1007

Popper, K.R.: Logic Lecture Notes 1939/41. Unpublished typescript (1939/1941). Karl Popper-Sammlung, Universität Klagenfurt, Box 366, Folder 19

Popper, K.R.: The Open Society and Its Enemies. Routledge, London (1945)

Popper, K.R.: Unpublished draft of a letter to Alexander Carr-Saunders (1946). Karl Popper-Sammlung, Universität Klagenfurt, Box 368, Folder 2

Popper, K.R.: Functional logic without axioms or primitive rules of inference. Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings of the Section of Sciences **50**, 1214–1224 (1947a)

Popper, K.R.: Letter to R. Carnap, November 24th 1947 (1947b). Karl Popper-Sammlung, Universität Klagenfurt, Box 282, Folder 24

Popper, K.R.: Logic without assumptions. Proc. Aristot. Soc. **47**, 251–292 (1947c)

Popper, K.R.: New Foundations for Logic. Mind **56**(223), 193–235 (1947d)

Popper, K.R.: Unpublished letter to P. Bernays, March 3rd 1947 (1947e). ETH Zürich, Paul Bernays Collection, Hs 975 : 3650. Translation by the authors

Popper, K.R.: Unpublished letter to P. Bernays, October 19th 1947 (1947f). ETH Zürich, Paul Bernays Collection, Hs 975 : 3651. Translation by the authors

Popper, K.R.: Corrections and additions to “new foundations for logic”. Mind **57**(225), 69–70 (1948a)

Popper, K.R.: On the theory of deduction, part I. Derivation and its generalizations. Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings of the Section of Sciences **51**, 173–183 (1948b)

Popper, K.R.: On the theory of deduction, part II. The definitions of classical and of intuitionist negation. Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings of the Section of Sciences **51**, 322–331 (1948c)

Popper, K.R.: Unpublished letter to P. Bernays, June 13th 1948 (1948d). ETH Zürich, Paul Bernays Collection, Hs 975 : 3654. Translation by the authors

Popper, K.R.: The trivialization of mathematical logic. In: Beth, E.W., Pos, H.J., Hollak, J.H.A. (eds.) Proceedings of the Tenth International Congress of Philosophy, vol. 1, pp. 722–727. North-Holland (1949)

Popper, K.R.: Part One: Autobiography of Karl Popper. In: Schilpp, P.A. (ed.) The Philosophy of Karl Popper, pp. 1–181. Open Court, La Salle (1974)

Popper, K.R.: Derivation and demonstration in propositional and functional logic (n.d.a). Karl Popper-Sammlung, Universität Klagenfurt, Box 36, Folder 20

Popper, K.R.: Table of contents for a textbook in logic (n.d.b). Karl Popper-Sammlung, Universität Klagenfurt, Box 371, Folder 1

Popper, K.R.: The propositional and functional logic of derivation and of demonstration (n.d.c). Karl Popper-Sammlung, Universität Klagenfurt, Box 36, Folder 21

Popper, K.R.: Unnamed typescript on the distinction between derivation and demonstration (n.d.d). Karl Popper-Sammlung, Universität Klagenfurt, Box 36, Folder 20

Popper, K.R., Bernays, P.: On systems of rules of inference (n.d). Karl Popper-Sammlung, Universität Klagenfurt, Box 36, Folder 13

Quine, W.V.O.: Mathematical Logic. W. W. Norton & Company, New York (1940)

Schroeder-Heister, P.: Popper's theory of deductive inference and the concept of a logical constant. Hist. Philos. Log. **5**, 79–110 (1984)

Schroeder-Heister, P.: Popper's structuralist theory of logic. In: Miller, D., Jarvie, I., Milford, K. (eds.) Karl Popper: A Centenary Assessment. Vol III: Science, pp. 17–36. Ashgate, Aldershot (2006)

Schroeder-Heister, P.: Proof-theoretic semantics. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy (2018). https://plato.stanford.edu/entries/proof-theoretic-semantics/

Seldin, J.P.: Curry's attitude towards Popper. In: Proceedings of the Annual Meeting of the Canadian Society for History and Philosophy of Mathematics, vol. 21, University of British Columbia, Vancouver, BC, 1–3 June 2008, pp. 142–149 (2008)

Whitehead, A.N., Russell, B.: Principia Mathematica. 2nd edn. Cambridge University Press, Cambridge (1925/1927)

David Binder holds degrees in philosophy, cognitive science and computer science and is currently pursuing a Ph.D. at the Department of Computer Science at the University of Tübingen, Germany. He has published on Karl Popper's work on duality and negation in logic, and is working together with Thomas Piecha and Peter Schroeder-Heister on an edition of Karl Popper's logical writings. His other work focuses on the propositions-as-types relation between logic and programming languages, and the duality between inductive and coinductive definitions.

Thomas Piecha is a researcher in logic in the Department of Computer Science at the University of Tübingen, Germany. He graduated both in Philosophy and in Physics, and holds a Ph.D. in Computer Science. He has published on proof-theoretic and dialogical approaches in logic, and currently works on constructive semantics and on an edition of Karl Popper's logical writings.

Logical Maximalism in the Empirical Sciences

Constantin C. Brîncuş

1 Popper's Logical Maximalism

Karl R. Popper (1947a, b) took the central topic of logic to be the theory of deductive inference and the main problem he was concerned with in his early writings on deductive logic was to give a satisfactory definition for the notion of "deductive valid inference". Although his definition was supposed to be a generalization of Tarski's definition of logical consequence, Popper showed that the notion of "truth" can be avoided, even though its use is not objectionable. In addition, unlike Tarski's model-theoretic (or, better, group-theoretic) criterion, he proposed an inferential criterion to draw a line between the formative (i.e., logical) and the non-formative signs (i.e., non-logical) and defined the validity of deductive inferences on the basis of inferential definitions.[1]

Although the notion of *truth* can be avoided in this inferential foundational approach to deductive inferences, in some latter writings Popper acknowledged that the validity of deduction goes beyond the signs and the rules that govern their use, and emphasized the constitutive role that the notion of *truth* has for logical deduction:

> Deduction, I contend, is not valid because we choose or decide to adopt its rules as a standard, or decree that they shall be accepted; rather, it is valid because it adopts, and incorporates, the rules by which truth is transmitted from (logically stronger) premises to (logically weaker) conclusions, and by which falsity is re-transmitted from conclusions to premises. (This re-transmission of falsity makes formal logic the *Organon of rational criticism* –that is, of refutation). (Popper 1962, 64)

In addition, Popper et al. (1970, 18) took the notion of truth to be crucial for the applications of logic in the other areas of inquiry. With regard to these applications,

C. C. Brîncuş (✉)
Institute for Research in the Humanities, The Research Institute of the University of Bucharest, Bucharest, Romania
e-mail: constantin.brincus@unibuc.ro

---

[1] For a detailed and extensive analysis of Popper's account on deductive inference and logical constants see Schroeder-Heister (1984).

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_9

he distinguished between two main uses of logic: a demonstrational use, for proofs, in the mathematical sciences, and a derivational use, for critical discussions, mainly in the empirical sciences. These two uses of logic mirror two essential model-theoretic features of logical consequence, namely, the transmission of truth from the premises to the conclusion (in proofs), and the retransmission of falsity from the conclusion to at least one of the premises (in critical discussions)—feature that actually makes logic *the organon of rational criticism*. This retransmission of falsity, modelled by *modus tollens*, places thus logic at the very centre of the methodology of the empirical sciences:

> The transmission of truth from premises to the conclusion means also the re-transmission of falsity from the conclusion to (at least one of) the premises. This is, from the pragmatic point of view, just as important an aspect of a valid deduction as the obtaining of reliable secondary information. It enables us to reject prejudices by falsifying their consequences; and it allows us to test a hypothesis by the method of trying to refute some of the conclusions which follow from it; for, if one of these is not true, the hypothesis cannot be true either. (Popper 1947a, 266)

The application of logic in the mathematical sciences and in critical discussions is supposed to be made in agreement to, what I will call, the requirements of *logical minimalism* and *logical maximalism*. Logical maximalism is the idea that in the derivational use of logic, in critical contexts, and in particular in the empirical sciences, one ought to use the strongest logic at our disposal. If we use a weaker logic, Popper contends that we are not critical enough:

> If we want to use logic in a critical context, then we should use a very strong logic, the strongest logic, so to speak, which is at our disposal; for we want our criticism to be severe. In order that criticism should be severe we must use the full apparatus; we must use all the guns we have. Every shot is important. It doesn't matter if we are over-critical: if we are, we shall be answered by counter-criticism. Thus we should (in the empirical sciences) use the full or classical or two-valued logic. If we do not use it but retreat into the use of some weaker logic - say, the intuitionist logic, or some three-valued logic (as Reichenbach suggested in connection with quantum theory) - then, I assert, we are not critical enough. (Popper et al. 1970, 18)

Conversely, logical minimalism is the idea that in the demonstrational use of logic, in mathematical proofs, one ought to use minimal logical means, such as intuitionistic, minimal or positive logic. So, in a mathematical proof it seems rational to attempt to minimize the spending of logical resources. Popper seems to regard logical minimalism as being actually embedded in the practice of the working mathematicians:

> Now, let us look, by contrast, at proofs. Every mathematician knows that considerable interest lies in proving a theorem with the help of *minimum apparatus*. A proof which uses stronger means than necessary is mathematically unsatisfactory, and it is always interesting to find the weakest assumptions or minimum means which have to be used in a proof. In other words, we want the proof not only to be sufficient, that is to say valid, but we want it if possible to be necessary, in the sense that a minimum of assumptions have been used in the proof. (Popper et al. 1970, 19)

Both logical minimalism and logical maximalism are two challenging assumptions. For instance, logical minimalism sounds more like a cost–benefit analysis of mathematical proofs. It aims at finding the minimal logical cost for proving a mathematical theorem. Certainly, it is not very clear whether Popper would prefer minimizing logical cost even if this incurred any mathematical or epistemological expenses. Logical minimalism by itself requires a detailed analysis, but my interest below, however, will be with its correlate, logical maximalism. The problem that I will be concerned with in the next section is whether the logical maximalism requirement makes the logic used in the empirical sciences immune to revision.

2 Is Logical Maximalism Compatible with a Revision of Logic?

If we consider logical maximalism strictly, with no other qualifications, it simply tells us that we *ought to* use the strongest logic at our disposal. In this case, if we already know which is the strongest logical theory on the theoretical market, then we have no other option left for a change of logic. Haack (1996, 37–38) describes Popper's motivation for logical maximalism, in relation to his falsificationist methodology, as follows:

> Popper's position seems to be like this: logic is a tool employed in the programme of attempted falsification. Since it is methodologically desirable that a test of an hypothesis should be as stringent as possible, the strongest possible logic should be used in deriving consequences from the hypothesis, so that its class of potentially falsifiers may be as inclusive as possible. This viewpoint is particularly forcibly expressed when Popper discusses the proposal that logic be modified in order to avoid certain 'anomalies' allegedly arising in quantum physics; if there are anomalies, Popper argues, they show that there is something wrong with quantum theory and modifying logic to avoid them is a dangerous evasion. [...] He wants to rule out altogether the possibility of *ever* resorting to change of logic rather of other beliefs.

Hence, according to Haack's interpretation, logical maximalism is a consequence of Popper's falsificationist methodology. Since we want to have empirical tests as severe as possible, then we should use the logic that allows us to deduce as many consequences as possible from the hypotheses, i.e., the strongest logic. If things are so, then a change in logic should not be preferred over a change in other beliefs. For instance, in the proposal of modifying logic in order to deal with certain 'anomalies' that seem to arise from quantum physics, logical maximalism advices us to look closely to the quantum theory itself and not to the logic used to derive from it consequences formulated in observational terms. For Popper, thus, the revision of logic seems to be contrary to progress in the empirical sciences. This is so because the weaker the logic, the less consequences one can deduce from hypotheses; the less consequences, the more chances to hold on to false hypotheses. By contrast, the stronger the logic, the more consequences one can deduce from hypotheses; the more consequences, the more chances to discard false hypotheses.[2] In this sense, strictly viewed, logic is not a proper part of science, since logical statements cannot be inconsistent with basic statements, that is, they are not scientific, i.e., falsifiable in the same way in which the empirical hypotheses are. However, as we shall see below, logic and physics are, in a certain sense, in the same boat for Popper.

It should be noted however that Popper's logical maximalism is a general requirement that applies to every rational dispute, not only to those from scientific contexts. The derivational use of logic is actually present in every critical discussion. Popper's main tenet is that disputing the logic that underlies a critical discussion may undermine the method of critical discussion itself. Thus, answering criticism by changing the logic is not, at least *prima facie*, an adequate attitude:

> What I should wish to assert is (1) that criticism is a most important methodological device; and (2) that if you answer criticism by saying, 'I do not like your logic: your logic may be all right for you, but I prefer a different logic, and according to my logic this criticism is not valid', then you may undermine the method of critical discussion. (Popper et al. 1970, 18)

One of the main objections that Haack raises to Popper's logical maximalism is that it would leave logic *totally immune* from criticism. So to say, we are not critical enough if we adopt logical maximalism; some parts of the web of knowledge will be accepted *dogmatically*, or only on the basis of a pragmatic justification through their usefulness in the process of critical discussions.

In order to see whether Popper had a non-critical attitude towards logic in the derivational contexts, it is fruitful to take a look at an exchange that he had with W.O.V. Quine on the relation between logic and physics at the *International Colloquium on Logic, Physical Reality, and History*, held in Denver in 1966. In his comments to Popper's paper 'A Realist View of Logic, Physics, and History', Quine agreed with many aspects embedded in Popper's view on logic (for instance, the non-conventionalism, the function of truth transmission and falsity re-transmission), but he also had some critical remarks in connection with the status of logic and physics in the overall web of knowledge:

> I am a logical realist, as Popper is, but I put logic and physics on the same level. They are truths about the world in the broadest sense of the word. But in the case of a disagreement

---

[2] Tennant (1985) takes the deductive structure of the *experimental refutation* of scientific hypotheses, present in Popper's analysis of scientific method, to have the following schema (P): from hypotheses and boundary conditions we obtain predictions which, together with observational reports, lead to contradiction. The role of logic in the Popperian methodology is taken by Tennant to consist in the retransmission of falsity from the contradiction encountered to at least one of the premises. After he proves that every classically inconsistent set of first order sentences (expressed in ~ , v, &, → , and ∃) is intuitionistically inconsistent and (expressed in ~ , v, & and ∃) is minimally inconsistent, Tennant concludes that both intuitionistic and minimal logic are adequate for *Popperian science*. I think that the reduction of the role of logic *only* to the downward direction in Schema P is not a good adviser for judging which logic is adequate for *Popperian science*. Logic is also needed for deriving the predictions from hypotheses and boundary conditions and this is an important reason for which we have to use *all the guns we have,* i.e., we have to use all the logical rules available in order to deduce consequences from the empirical hypotheses. From this point of view, classical logic has an advantage, namely, it provides us with more logical tools than the other systems of logic.

> that is so fundamental as a disagreement on logical principles, the only thing we can do is talk about talking. But this, again, isn't peculiar to logic and it doesn't make logic a matter of convention. The same thing happens in physics. The same thing happens whenever an issue takes the form of disagreement over extremely fundamental matters, matters so fundamental to the conceptual scheme that if you walk at that level you keep begging the question by simply saying, "Well, it is obvious from what you said that this follows." This happens not only in logic. There is something of this kind when we start talking about relativity theory or quantum mechanics. One has to stand off, talk about the system, and appeal to the pragmatic value of this or that system. One says: Here is a simpler formulation that takes in all the data that the old formulations took in, and, moreover, does it more simply. (In Popper et al. 1970, 30–31)

Quine's main point is that Popper's "view about the function of logic draws no clear boundary between logic and other sciences". For instance, in an argument with three premises and a conclusion, logic is used to transmit truth from premises to the conclusion, but this argument could be also rewritten such that it has two premises and the third one is absorbed as the antecedent of a material conditional in the new conclusion. This material conditional obtained in actually a sentence of "physics or biology, or whatever the subject matter may have been". Quine also points out that the properties of logical consequence of transmitting truth and retransmitting falsity do not provide a criterion of delineating the province of logic. Moreover, if a disagreement on fundamental matters occurs, the main thing that has to be primarily done is critical discussion. But this is common both in logic and in physics. This is one of the reasons for which Quine takes logic and physics to be in the same boat. The web of beliefs confronts experience as a whole and a conflict with experience may bring changes even in the central parts of the web. Logic is thus also exposed to revision, as the physical hypotheses are.

In his reply to Quine, Popper admitted that, up to a point, logic and physics are in the same boat, in the sense that both of them are criticizable. Nevertheless, he immediately recognized that a weakening of classical logic, as suggested by Birkhoff and von Neumann, or by Reichenbach, is not an adequate response for the difficulties from quantum physics:

> There seems to be complete - or almost complete - agreement between Quine and myself. Where I thought that we disagreed was on Birkhoff and von Neumann's "Logic of Quantum Mechanics": I thought (perhaps mistakenly) that he had taken this famous paper to show that logic and physics are in the same boat. Up to a point no doubt they are: both are criticizable. But a mere weakening of classical (or Boolean) logic, as suggested by Birkhoff and von Neumann, and also by Reichenbach, does not seem to me an adequate response to a difficulty in one of the empirical sciences. It is a strategy which we ought to exclude if we want to learn from experience, because every sufficient weakening of the underlying logic will make any empirical theory for ever secure. It will make it uncriticizable, irrefutable. These remarks show that in some cases at least we can critically discuss a proposed change in logical theory. (Popper et al. 1970, 35)

Popper's reply is made in accordance with the logical maximalism requirement. If we weaken the underlying logic, then we are in danger of being insufficiently critical with the scientific hypotheses. A sufficient weakening will make the scientific hypotheses uncriticizable and, thus, irrefutable.[3] We can easily understand that there is a tension between the criticizability of logic and the criticizability of empirical hypotheses. If one increases, the other one decreases. Certainly, the strongest logic appears indeed to be the one which allows the maximum criticizability of empirical hypotheses, which is the main target of the falsificationist methodology. However, since Popper admits that both logic and physics are criticizable, this means that logic is not taken dogmatically to be immune to revision. Although Popper does not develop in details the sense in which logic itself is criticizable in its derivational use, some useful remarks on this matter are made by Miller (1994, 90–91):

> In addition to the point *a* above, that logical truths are not criticizable, there is much more serious point concerning the uncriticizability of our system of rules of logical inference. These are not, like tautologies, mere consequences of other positions we adopt but vital constituents of the rationalist position itself. […] Critical argument certainly cannot be carried on without some system of logic. You cannot in this sense abandon logic and remain a rationalist. […] But the system of logic employed can –despite Lewis Carroll- be taken as one of the things under investigation in a critical discussion. […] The rules, after all, are universal, and if they were systematically to lead to error in a way that in fact they do not, we might eventually decide that we have got them wrong. We must not be misled by the fact that we have not got them wrong.

Miller operates the distinction between logical truths and rules of logical inference. The rules, of course, are very important for the critical rationalist, being central in the process of critical argumentation. As Tarski (1941/1995, 47) emphasized, "the rules of inference, which must not be mistaken for logical laws, amount to directions as to how sentences already known as true may be transformed so as to yield new true sentences." The rules, thus, allow us to see explicitly the consequences of the beliefs that we hold and to evaluate them accordingly. Miller recognizes that we can evaluate in a critical discussion even the rules of inference themselves. Being universal sentences, any *counterexample* would suffice to invalidate them. However, it is not very clear in what a *counterexample* would consist of.[4] Their revision would be motivated "if they were systematically to lead to error *in a way that in fact they do not*" (my emphasis). Once again, the requirement of logical maximalism is not theoretically incompatible with the revision of logic, although it is not quite clear what kind of evidence would make us to revise the system of logic used in critical discussions. The expression "in a way that in fact they do not" seems to suggest that,

---

[3] Mortensen and Burgess (1989, 48–49), in reply to Popper et al. (1970), have proposed a sort of logical minimalism in the empirical sciences, i.e., "prefer a weaker logic". Their main reason for this principle is that a weaker logic has a larger number of theories and a theory criticized from a weaker logical base excludes the option of modifying it to enter in dispute with those of a stronger logic. This idea, however, in a different shape, is still debated nowadays. Williamson (2017, 337) briefly dismisses the idea that weakness is a strength in logic, since almost every logical principles has been subjected to criticism. Thus, a weaker logic is not necessarily in a better position. Bell (2019, 213–217) offers a detailed analysis of this idea, but he concludes that Williamson's argument for logical strength does not work, unless we assume a sort of Quinean conservatism principle (see also footnote 10 below).

[4] This problem will be discussed in the last section of this paper.

at least until present, the system of logic used in critical discussions, presumably the classical one, is not open to refutation.

To sum up the above discussion, logical maximalism tells us that we *ought to* use the strongest logic at our disposal in critical discussions, but if we *ought to* use the strongest logic in these contexts, then logic seems to remain *immune* to criticism. However, the critical rationalists[5] admit that we can critically discuss a change in logical theory. Therefore, the logical maximalism requirement can be itself subjected to criticism. The question that arises at this point is the following: what kind of *evidence* would lead the critical rationalist to revise the system of logic that underlies a physical theory, such as quantum physics. Is it the same kind of evidence as in the case of the empirical theories or it is more like a pragmatic decision of changing the conceptual framework in which the empirical theories are formulated? A better understanding of this question will be gained once we analyze the abductivist view on logic and compare Popper's critical rationalism with it.

3 The Abductivist Assessment to Logic

It is a widespread conviction nowadays among logicians and philosophers that logic should be assessed abductively, i.e., on the basis of empirical evidence and theoretical virtues. The abductive methodology is a rational reconstruction of the way in which the scientists select and justify their hypotheses and theories. In its modern use, abduction is better known as the inference to the best explanation. Roughly expressed, a scientific theory is selected as the best one, according to the abductivists, on the basis of its theoretical features and of its explanatory and predictive power. For instance, Williamson (2017, 334) synthesizes the abductive methodology as follows:

> We make the standard assumption that scientific theory choice follows a broadly abductive methodology. Scientific theories are compared with respect to how well they fit the evidence, of course, but also with respect to virtues such as strength, simplicity, elegance, and unifying power.

Williamson takes this methodology to be 'the best one science provides' and for this reason it should also be applied in selecting a logical theory as being the best one.[6] He does not provide us with many clarifications as to what would constitute evidence for a logical theory, but he gives us, however, a negative determination, namely: 'Evidence is not confined to observations. We may use anything we know

---

[5] Critical rationalism is defined by Miller (2012, 93) as "the generalization, from the empirical sciences to the whole of our knowledge, of the methodological falsificationism (or deductivism) proposed by Karl Popper in *Logic der Forschung*".

[6] The abductivist approach to logic is characterized by Russell (2019, 550) as consisting in two main claims about the *justification* of logic. The first one is the holism about justification, namely, that it is entire logics that are subject to justification, and not particular claims of logical consequence—logic being taken as a theory of the relation of logical consequence. The second one is that the justification is given by adequacy to the data, and the possession of virtues and absence of vices.

as evidence.' Williamson (2017, 335). That *evidence* is not restricted to observations means that it can also incorporate our intuitions in certain contexts, as the modal ones. Nevertheless, these remarks on *evidence* are very general and could be accommodated with different philosophical views. For instance, we may understand them in Quine (1951)'s sense, namely, that certain empirical observations might determine us to revise even some sentences situated in the central parts of the web, as the logical principles are. But we may also interpret them in Carnap (1950)'s sense, namely, that some empirical observations could suggest us to change the conceptual framework we work in and thus the logic constitutive of that framework. Nevertheless, letting these interpretations aside, it is a clear point that for the abductivist the *empirical* evidence is relevant for choosing or changing a logical theory. Certainly, it remains the same question as for the critical rationalists: in what way empirical evidence may determine us to change a logical theory already at use in a certain area of inquiry?

Beside evidence, the other criteria that are used by the abductivist to choose a logical theory are the theoretical virtues. One of the most important virtues is taken by Williamson (2017) to be *strict logical strength.*[7] In this sense, a logical theory T is stronger than T* if and only if every theorem of T* is a theorem of T, but not conversely. Since two internally consistent theories may be inconsistent with each other, Williamson (2017, 336) introduces a looser sense of strength, i.e., *looser scientific strength.* In this looser sense, a theory T is stronger than a theory T*, with which it may be externally inconsistent, on the ground that 'the former is more specific or informative than the latter'. Both senses have applications in his view to logical theories. Williamson claims that if a theory T is stronger that T* in the strict sense, then is also stronger in the looser sense, but not conversely.[8] In addition, and this is the point in which he and Popper converge, Williamson (2017, 337, 338) emphasizes that logical strength is one of the virtues that makes classical logic the most adequate instrument in inquiry:

---

[7] Russell (2019, 556) argues that logical strength is theoretically neutral in the selection process of a logical theory, i.e., it is neither a virtue nor a vice. However, Russell discusses the problem outside a specific context and thus she concludes that "logical strength is something that a logic is supposed to *get right*, rather than something that it is always good to have more of". Roughly expressed, the idea is that for certain purposes a weaker logic may be better than a stronger one. Certainly, Popper would agree with this idea, since he actually emphasized that, in the demonstrational use of logic, "the weaker the logical means we use, the less is the danger of consistency" (Popper et al. 1970, 20). Nevertheless, it is worth mentioning that the distinction between the demonstrational use of logic and its derivational use probably does not make much sense for the abductivists, who work, so to say, in a Quinean holist framework.

[8] In reply to Williamson, Russell (2019, 557–562) argues that logical strength does not entail scientific strength since, among other reasons, weaker and stronger logics are on a par if informativeness refers to the classification of arguments in valid or invalid. Certainly, this sense of informativeness is *internal* to the province of a certain logical theory, i.e., the validity of an argument is relative to a certain logical theory. However, as Williamson (2017, 337) exemplifies, if we extend the classical and the intuitionistic propositional calculi to include quantification into sentence position, ⌜(∀p)(p v ~ p)⌝ will be a theorem in the extended PC, while ⌜ ~ (∀p)(p v ~ p)⌝ will be assertable in the extended IC. Since a universal generalization is more informative than its negation, there is at least one clear sense in which a logical theory is more informative than another. (For an elaborate discussion on the relation between logical and scientific strength see Incurvati and Nicolai (2020)).

> Once we assess logics abductively, it is obvious that classical logic has a head start on its rivals, none of which can match its combination of simplicity and strength. Its strength is particularly clear in propositional logic, since PC is Post-complete, in the sense that the only consequence relation properly extending the classical one is trivial (everything follows from anything). First-order classical logic is not Post-complete, but is still significantly stronger than its rivals, at least in the looser scientific sense, as well as being simpler than they are; likewise for natural extensions of it to more expressive languages. In many cases, it is unclear what abductive gains are supposed to compensate us for the loss of strength involved in the proposed restriction of classical logic.

Although Williamson prefers classical logic and its extensions, in particular higher-order modal classical logic, he lets open the possibility that some non-classical logic "can overcome the initial advantages of classical logic once we move to a wider setting, by considering fit with evidence or with other scientific theories, or by treating more expressions as logical constants" Williamson (2017, 338). The paradigmatic historical case where the use of classical logic in the empirical sciences was challenged is the proposal of introducing quantum logic in order to simplify and get a better understanding of quantum phenomena. The point of view of the abductivist approach is to compare the theory generated from the principles of quantum mechanics by using classical logic and the theory generated from them by using quantum logic. Relying on Quine (1970, 85–86)'s and Putnam (2012)'s analyses, Williamson (2017, 338) concludes that 'in practice, rejecting classical logic just does not seem to help us understanding the nature of quantum reality'.

Williamson's attitude towards classical logic is not being based primarily just on strength and conservatism.[9] He considers that it 'has been tested *far* more severely than any other logic in the history of science, most notably in the history of mathematics, and has withstood the tests remarkably well' Williamson (2017, 338). His position seems to converge here to Miller (1994)'s view expressed above, namely, that logical rules would be changed if they were systematically to lead to error in a way that in fact they do not, since the tests to which they have been subjected thus far have corroborated them.[10]

---

[9] See Bell (2019) for a very good analysis and criticism of Williamson (2017). He argues, however, that the conservatism principle, or something similar to it, underlies "Williamson's new Quinean argument".

[10] The severe tests that Williamson refers to are interpreted by Bell (2019, 220) in the sense that the theories which have successfully passed the testing process appear to be closed under the *classical* relation of logical consequence. However, Bell does not consider this fact as a disadvantage of weaker logical theories, since a theory closed under classical logic is also closed under a sub-classical one. Moreover, the successful testing of classical logic in the history of mathematics appears so since in mathematics the theoretical possibilities are usually reduced in practice to classical models. Certainly, Popper would agree with Bell here, since, in the demonstrational use of logic, using weaker means is a better option.

4 Final Remarks: Critical Rationalism Versus Abductivism

We have seen thus far that the critical rationalism methodology and the abductivist one (as it is described and used by Williamson) lead to the same result concerning the selection of a logical theory for being used in the empirical sciences[11]: classical logic is the best candidate for doing scientific research. Although both Popper and Williamson consider that the condition of logical maximalism is not an absolute one and that the classical logic could itself be replaced, they are not very informative as to what would constitute a clear case of evidence for changing a logical theory.

Despite this fact, however, there is a fundamental point in which the two methodologies go apart. The critical rationalists have a neutral attitude towards theoretical constructs (concepts, ideas, theories, conceptions), namely, they do not look for justifications for them, but rather they do their best in order to criticize them as much as possible. To look for justifications for a certain idea means to already have a certain *attachment*[12] towards it, which may incline you to interpret as evidence something that actually might not be so. For instance, if you already have a certain attachment towards classical logic (since you are very familiar with it, you taught it, you worked a lot with it), then you will be inclined to see it confirmed by a lot of things, in a lot of cases. In contrast, if you simply accept it because it resisted to the criticisms to which it has been exposed so far, and remain open to further criticisms that may, *in principle*, overthrown it, then you have a more adequate theoretical and practical attitude with regard to the powers and the results of our reason. The distinction between justification and criticism was very nicely emphasized by Miller (2012, 95):

> Central to the philosophy of rational criticism is the realization that the process of reasoning can never provide justification, but it may provide criticism; and indeed, that the rational attitude consists wholly of openness to criticism, and of appropriate responses to criticism. Justification, conclusive or inconclusive, is revealed as neither possible, nor useful, nor necessary.

Consequently, from the perspective of critical rationalism, anything can count as evidence for revising a logical theory. Whatever constitutes itself in a strong criticism for revising a theoretical construct will do the job. The only thing that the logical maximalism requirement imposes, in my understanding, is a certain order in the processes of testing and revision, namely: we should not try *prima facie* to revise the logical statements, since this action may have enormous consequences in the total system of science, due to their central role in the system and to the logical interconnections among the other statements of the system. What we should rather do

---

[11] Williamson does not seem to make a clear-cut distinction between the use of logic in the empirical sciences and in the mathematical ones, a thing quite natural if we consider the Quinean background of the abductive methodology. Nevertheless, since he selects classical logic as the best candidate *simpliciter*, its use in the empirical sciences is also presupposed here.

[12] By this I mean both epistemic and psychological attachment. Our ideas are embedded in the overall web of our mind and are important tools for the accommodation in natural and social environment. Since they are *ours*, we do not have all the time an objective and critical attitude towards them.

is to use the strongest logic in order to identify the problematic parts of the scientific theories.

The abductivist approach to logic, being focused on justification, I think that is exposed to have a less critical attitude to logic. In my understanding, justification is just a particular instance of criticism. By this I mean that we label a certain idea to be justified if it resisted, thus far, to the criticisms to which it has been subjected. Someone may say that this is just a terminological issue. However, I think that it is not just so because the function of the justification concept in the economy of our thinking induces to us the tendency of being attached to the beliefs for which we think we posses *justification*—thing that makes us less receptive to good criticisms that may be *prima facie* ignored due to this attachment. Certainly, this is a psychological aspect, but it should not be neglected.

Another important point that is relevant for the understanding of the revision process of logic in the derivational contexts was made by Popper in his contribution to the symposium *Why Are the Calculuses of Logic and Arithmetic Applicable to Reality?*. Popper (1946, 42–44) analyzed the distinction between rules of inference and formulas of logical calculi in connection with their applicability to, what he called, *reality*. In his view, logical rules of inference should not be confused with the sentences of a logical calculus. The former are about sentences, i.e., are meta-sentences, and assert unconditionally that every sentence of a certain form is inferable from a set of sentences of another form, while the latter are about the semantic values of the sentences and the relations among them. The same rules of inference may be used for constructing different logical calculi (for instance *modus ponens*).

Popper (1946, 48) defines a rule of inference such that it cannot have a counterexample. It always leads from true premises to a true conclusion: "we call a rule of inference 'valid' if, and only if, no counterexample to this rule exists; and we may be able to establish that none exists". In this sense, by definition, a rule of inference cannot be refuted.[13] So to say, it is *analytic*. Yet, he accepts that we can have alternative logics by formulating "alternative rules of inference with respect to more or less different languages" Popper (1946, 51). Popper's assumption here is that the rules of inference are '*to a certain extent,* always relative to a language system'. Consequently, it seems that for Popper, as for Carnap and Quine, to say that a specific rule of inference is invalid simply amounts at changing the *subject,* that is, changing the conceptual scheme in which we work, scheme that provide the logical constants with their actual meaning at use.

Regarding the logical calculi, Popper has a quite similar attitude. Logical calculi[14] are constructed on the basis of the rules of inference from primitive formulas. In a

---

[13] To have a clear cut understanding of this idea we may think to the decision procedures in a logical system. Think for instance to the method or normal truth tables in classical propositional logic. We can easily decide I this case that there is no *permissible* possibility to have true premises and a false conclusion for modus ponens once we precisely disambiguate all the premises and the conclusion and accept the principle of bivalence. (See also Miller (2012, 100–103) for a discussion of the status of the logical rules of inference from a critical rationalist perspective).

[14] Popper takes a logical calculus to be either a derivational logic or a demonstrational one. The former is a system of logic "intended from the start to be a theory of inference in the sense that

way that resemblances quite well Einstein (1921, 233)'s view on the relation between geometry and experience,[15] Popper (1946, 54) concluded that the relation between logical calculi and *reality* has the following dynamics:

> in so far as a calculus is applied to reality, it loses the character of a *logical* calculus and becomes a descriptive theory *which may be empirically refutable*; and in so far as it is treated as irrefutable, i.e., as a system of *logically true* formulae, rather than a descriptive scientific theory, it is not applied to reality.

We see thus that, as in the case of the rules of inference, logical formulas are taken to be true by definition. In this sense, as far as the calculus to which they belong is considered *in itself*, we have no counterexamples to them. Once the calculus is applied, i.e., it is used in a certain descriptive theory (physical, for instance), it becomes subject of philosophical criticism. The formulas of the calculus are treated now as hypotheses. However, the logical maximalist constraint advices us to criticize them only after we have used all the *shots* these formulas provide us for criticizing the empirical hypotheses.

All these being said, it seems to me that Popper's logical maximalism requirement is an adequate rational attitude, both theoretically and practically, on the use of logic in critical discussions, since it advices us to derive as many consequences as possible from a given hypothesis or theory in order to confront them with experience through observations and experiments. Once this process has been performed, if problems still persist, of course, we can critically discuss the logical instrument at use. The process is here, so to say, dialectical. More precisely, it proceeds by the trial and error method.[16] By this I mean that we start to criticize the logical rules of inference or the logical formulas[17] that seem more problematic (as are the law of distributivity or the law of excluded middle in the case of quantum phenomena) and we keep using the other rules of inference and logical formulas. In the long run, the output of the criticism process may reside in changing some parts of the mathematical formalism

---

it allows us to derive from certain informative (non-logical) statements other informative statements" Popper (1947b, 230). This logic contains rules of inference for drawing consequences from hypotheses. A system of natural deduction fits very well this role. In contrast, "most systems of modern logic are not purely derivational, and some (for example, in the case of Hilbert-Ackerman) are not derivational at all." Popper (1947b, 230). These systems are demonstrational logics and are serving better in the demonstrational use of logic. The derivations conducted in them usually start from logical axioms, definitions or theorems. It should be noted however that this distinction refers primarily to the deductive format of the system and not to its strength.

[15] "As far as the propositions of mathematics refer to reality, they are not certain; and as far as they' are certain, they do not refer to reality".

[16] We apply the trial and error method whenever we are faced with a problem, we tentatively advance a solution, and then criticize it as much as possible in order to eliminate the possible errors. A detailed analysis between the dialectic and the trial and error methods is offered by Popper (1940). As he emphasized, the dialectic development could be actually explained on the basis of the trial and error method, which has a wider application.

[17] Certainly, if the deduction theorem holds for the system of logic we use, this distinction is less relevant.

of the physical theory, in providing new interpretations for the formalism,[18] or in abandoning certain logical principles. If we look at the quantum mechanics case, it seems that, thus far, changing the logical theory was not a fruitful path to follow.[19]

Acknowledgements I would like to thank my audiences in Klagenfurt and Prague, where parts of my work on this paper have been presented, in particular Bernard Burgoyne, David Miller, Jagdish Hattiangadi, Olival Freire, Flavio Del Santo, and Stephen Senn. I also want to thank Mircea Flonta and Iulian Toader for helpful discussions. This work was supported by a grant of the Romanian Ministry of Education and Research, CNCS - UEFISCDI, project number PN-III-P1-1.1-PD-2019-0901, within PNCDI III.

References

Bell, Jc.: On Williamson's new Quinean argument against nonclassical logic. Aust. J. Log. **16**(7), 202–230 (2019)

Brîncuş, C.C.: Philosophical accounts of first-order logical truths. Acta Anal. **34**(3), 369–383 (2019)

Carnap, R.: Empiricism, semantics and ontology. Rev. Int. Philos. **4**(2), 20–40 (1950)

Einstein, A.: Geometry and experience (1921). In: Einstein, A. (ed.) Ideas and Opinions (fifth printing), pp. 232–246, Crown Publishers, New York, (1960)

Haack, S.: Deviant logic, fuzzy logic: beyond the formalism. The University of Chicago Press (1996)

Incurvati, L., Nicolai, C.: On logical and scientific strength. Unpublished manuscript. https://philarchive.org/archive/INCOLA Accessed 27 July 2020

Miller, D.: Critical rationalism: a restatement and defence. Open Court, Chicago (1994)

Miller, D.: Overcoming the justificationist Addiction. Studia Philosophica Wratislaviensia: Supplementary Volume, 93–103 (2012)

Mortensen, C., Burgess, T.: On logical strength and weakness. Hist. Philos. Log. **10**, 47–51 (1989)

Popper, K.R.: What is dialectic?. Mind **XLIX**(194), 403–426 (1940)

Popper, K.R.: Why are the calculuses of logic and arithmetic applicable to reality?. In: Ryle, G., Lewy, C., Popper, K.R. (eds.) Symposium: Why Are the Calculuses of Logic and Arithmetic Applicable to Reality?. Aristotelian Society Supplementary, vol. 20, no. 1, pp. 40–60 (1946)

Popper, K.R.: Logic without assumptions. Proc. Aristot. Soc. **47**, 251–292 (1947a)

Popper, K.R.: New foundations for logic. Mind **56** (223), 193–235 (1947b)

Popper, K.R.: Science: conjectures and refutations. In: Popper, K.R. (ed.) Conjectures and Refutations: The Growth of Scientific Knowledge, pp. 32–65, Basic Books, London, Routledge (1962)

Popper, K.R.: A realist view of logic, physics, and history. In: Bondi, H. (ed.) Physics, Logic, and History, pp. 1–37. Plenum Press, New York (1970)

---

[18] It is important to mention that Birkhoff and von Neumann's proposal, and likewise Reichenbach's proposal, of introducing a new logic to deal with quantum phenomena was not for *testing* the quantum mechanics, but rather for obtaining a better understanding of it. See Putnam (2012) for a captivating story of the philosophical development of quantum logic.

[19] In addition, there are also some general considerations which suggest that logic cannot be treated in the same manner in which the physical hypotheses are, although both of them are criticizable. For a philosophical discussion on the status of classical logic, viewed from a metalogical point of view, see Brîncuş (2019).

Putnam, H.: The curious story of quantum logic. In Putnam, H. (ed.) Philosophy in an Age of Science: Physics, Mathematics, and Skepticism, ed. M. De Caro and D. Macarthur, pp. 162–177, Cambridge, MA: Harvard University Press (2012)

Quine, W.O.V.: Two dogmas of empiricism. Philosophical Review **60**(1), 20–43 (1951)

Quine, W.O.V.: Philosophy of logic. Prentice Hall, Englewood Cliffs, NJ (1970)

Russell, G.: Deviance and vice: strength as a theoretical virtue in the epistemology of logic. Philos. Phenomenol. Res. **XCXI**(3), 548–563 (2019)

Schroeder-Heister, P.: Popper's theory of deductive inference and the concept of logical constant. Hist. Philos. Log. **5**, 79–110 (1984)

Tarski, A.: Introduction to logic and to the methodology of the deductive sciences (1941). Dover Publications, New York, Translated by Olaf Helmer (1995)

Tennant, N.: Minimal logic is adequate for Popperian science. Br. J. Philos. Sci. **36**(3), 325–329 (1985)

Williamson, T.: Semantic paradoxes and abductive methodology. In: Armour-Garb, B. (ed.) Reflections on the Liar, pp. 325–346, Oxford University Press (2017)

Constantin C. Brîncuş is Researcher at the University of Bucharest, Romania. His research interests are mainly in logic and its philosophy, epistemology, philosophy of mathematics, general philosophy of science and value theory. Recent articles include "Are the Open-Ended Rules for Negation Categorical?" in Synthese (2019), "Categoricity and Negation. A Note on Kripke'sAffirmativism" in Logica Yearbook 2018, and "Philosophical Accounts of First-Order Logical Truths" in Acta Analitica (2019). E-mail: constantin.brincus@unibuc.ro

The Role of Logic in Science

Nimrod Bar-Am

Logic is the science of proof. But, proof of what? "…since people have tried to prove obvious propositions they have found that many of them are false", noted Russell astutely (Russell 1963, 61). Historical answers to our question are, therefore, quite diverse, despite what anachronistic studies of logic may imply. They comprise the long and complex story of the development of logic from Aristotle to Gödel, reflecting the time-old rear action battle fought by classic epistemologists against skepticism. Popper's reconstruction of logic as the theory of refutation, has helped us realize that the skeptic has effectively won this historical battle, and that securing empirical knowledge by logical means alone (be they inductive, analytic or transcendental) is a futile effort. And it helped us move forward to the cleaner concept of logic as expressing the most basic methodological procedures accepted in science. As such, disagreements between logicians are metaphysical, or heuristic controversies about the proper (fruitful and convenient) methodological rules for conducting science.

Even before Aristotle's revolutionary invention of the logical variable philosophers did their best to reason convincingly and even quite systematically, and of course they sometimes succeeded. Parmenides used what we nowadays would call analytical reasoning in an attempt to prove his incredible theory, and he titled it "*piston logon*" (proven assertion). Zeno, his student, has provided us with rough yet brilliant versions of *reductio ad absurdum* of the opposite theory (the Democritian claim that the cosmos contains a void, and that it, therefore, allows for motion and time). Famously, and quite brilliantly, the sophist Gorgias ridicules the pretentiousness of their argumentation style by demonstrating that it can easily be applied to Parmenides' own theory, thus yielding its refutation as well as the refutation of its negation (an exercise that Plato himself repeats in a enigmatic manner, without

---

N. Bar-Am (✉)
Communication Department, Sapir College, Shaar Hanegev, Israel
e-mail: ngbaram@gmail.com

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_10

discussing its meaning, in his “Parmenides”). Gorgias thus exposed ancient epistemology’s dead-end. Socrates heroically refused to allow such defeatism. Although he typically admits ignorance in matters epistemological he nevertheless recommends relentlessly attempting to refute the knowledge claims of his interlocutors, declaring successful refutations to be genuine epistemic achievements. Having no logical variables to his disposal he utilizes a repetitive style known as Socratic *elenchus* which, if it was abstracted and generalized, would have amounted to what we nowadays call *modus tollens*. Finally, participants of Plato’s later dialogues (notably “The Sophist” and “The Statesman”) rekindle Eleatic pretentiousness, as they seek to establish definitions of various subject matters by a procedure of logical analysis known as *diaeresis*. If this procedure were to be generalized, it would have provided us with the following logical rule: x is either a or b; it is not a, therefore it is b.

In this lively environment of the pre-history of logic brilliant philosophers search and test various argumentation styles and semi-explicit debate procedures that will later on be united under the study of Logic. Three very distinct approaches stand out here, and they are worth mentioning because they set the stage for all future controversies about the status of logic. The first, which we can call “sophistic” and even “relativistic”, declares all argumentation procedures epistemically barren albeit (sometimes) rhetorically effective: they are, as Gorgias had claimed, mere manifestations of rhetorical wizardry, nothing more. The second approach, which we can call “reluctantly skeptic”, declares logical enquires into the nature of reality a worthy heroic effort. It stresses that refutation of a theory is a genuine epistemological achievement, as it validly demonstrates its falsity. But it forbids us to derive from it the absolute truth of its negation. For a refutation too can, one day, be refuted. As Socrates has put it: the oracle declaring him the wisest of all men must be right, but only because he (Socrates) is aware of his own ignorance. The third approach towards these argumentation procedures should be titled “dogmatic”. It seeks to establish incontestable foundations for science, and then validly reason in an attempt to derive the rest of knowledge from them. This attitude is implied by the Eleatics and by the mature Plato, and was adopted, developed and systematized by Aristotle . Its influence upon the history of logic cannot be underestimated, as it shaped its relationship with science and metaphysics for over two millennia, until Frege and Russell, Tarski and Popper had altogether changed our view of the matter, ridding logic from the impossible burden of proving empirical science, returning us to a more Socratic point of view.

For in Aristotle it is clear that dialectics is not merely a mode of argumentation that allows one to carefully formulate informative theoretical conjectures, but also a mean to establish them as uncontestable basic truths, aka ‘essential definitions’, which will function as premises for scientific (apodeictic) syllogisms. This applies equally for inductive syllogisms, in Aristotle, for Aristotle regards induction as a type of dialectical argument (*An. Pos. 71a4*). Now, since today we take it as a matter of course that this task—logically proving informative theoretical knowledge—is unfeasible, we must be very careful when inspecting the situation: we must not allow our utmost respect for Aristotle to blur our realization that he has committed what nowadays we regard as a highly influential philosophical error. Indeed respect for Aristotle is typically so great that it is remarkably rare to find in the learned literature attempts to reconstruct his error. I have endeavored to analyze it in great detail (Bar-Am 2008). Let me try and briefly sum up things for you here.

Both Plato and Aristotle were greatly impressed by the fact that seemingly immediate and incontestable observations presuppose, or at least seem to presuppose, a great deal of theoretical knowledge of universals and their taxonomical hierarchies, which we perhaps were unaware of at first, but that we can easily become aware of by critical, logical analysis, by dialectics and/or induction (*An. Pos. 71a6*). For example, the observation "this rose is red", which seems so immediate as to be incontestable when facing a red rose, seems to presuppose the taxonomical knowledge "Red is a Color", which subtly conflates empirical knowledge (the existence of red things in our cosmos is of course a contingent fact) and an analytical appearance (clearly, red is a color, as every English speaker knows). Similarly, (or, more accurately, misleadingly similarly) the famous dialectical inquiry performed by Socrates and the slave in Meno, based on a drawing in the sand, leads the slave (and us readers) to recognize that we were implicitly all along in possession of an apriori notion of a (semi-general) case of the Pythagorean theorem.

The general epistemological idea behind these examples (which is shared by Plato and Aristotle) is that there exists a grand matrix of universals, of natural kinds, that orders them according to their proper taxonomical relations, which is somehow presupposed by our observations, and which is obtainable, exposable and extractable, by logical analysis, by dialectics and induction. This grand matrix of universals, Aristotle argues, is empirical science in its entirety. Logic, for him, is therefore not merely the method of expressing the grand matrix by means of apodeictic syllogisms, but also the method of exposing and establishing it by means of dialectical and inductive syllogisms. This point is central: the isomorphism between the method of expressing the taxonomical relations that science comprises (apodeictic syllogisms) and the method of exposing and establishing them as essential definition (dialectical and inductive syllogisms) is the heart of Aristotle's epistemology, indeed it is so central to it, that it features the opening remarks of his Prior Analytics (*An. Pr. 24a23–24b13*), the opening remarks of his Posterior Analytics (*An. Pos. 71a4-9*), as well as the opening remarks to his Topics (*Top. 100a25–100b24*). It is the birth of the myth that the skeptic can be answered by logical means alone, for we can somehow, to use Aristotle's own words, "prove the universal from the self evident nature of the particular" (*An. Pos. 71a6*).

Aristotle's theory that the grand matrix of science is extractable by logical means alone, (from our observations by induction, and from our critical inquiries by dialectics, that is by a logical analysis of our concepts) is perhaps the most influential epistemology ever formulated. It is also very vague. For Aristotle never made it clear how exactly the process is to be performed and why it guarantees the obtainment of empirical truth. Clearly, induction may lead us astray: we speculated that all swans are white until Tasmania was discovered. And just as clearly, if you and I conclude a dialectical conceptual inquiry with the conclusion that absolute speed cannot exist (as Leibniz had done), this does not make it into an empirical fact, as Einstein had shown. Induction and logical analysis may, perhaps, be excellent tools for formulating conjectures, but they are no tools for proving them. And so, the greatest minds in the history of philosophy have endeavoured to break up and reconstruct the missing pieces in Aristotle's claim that science can be proven on logical reasoning alone, or at most, on logical reasoning and uncontestable immediate experiences.

Consider Leibniz for example. Greatly impressed by the Aristotelian idea that logical analysis of concepts may lead to the intuitive recognition of essential definitions, he sharpened Aristotle's rough notion of proof in an attempt to improve and complete Aristotle's program. He explicitly suggested that, essentially, all empirical truths are analytic, and that demonstrating that they are analytic is tantamount to proving them. Perhaps, he added, we find it difficult to currently realize all this because our current conceptual framework is not yet isomorphic to the grand matrix of being, the cosmic taxonomy of natural kinds, but should we succeed in constructing the perfect language, a semantic framework that would perfectly correspond to the grand matrix of universals, we would be able to prove that all empirical truths are analytic. He left us dozens of drafts that are supposed to detail how such a language would look like. But of course he never constructed one.

Or consider Kant. Overwhelmed by Hume's (rather trivial) observation that inductive arguments are not really isomorphic to apodeictic syllogisms, indeed that, strictly speaking, they are invalid inferences, he was nevertheless greatly impressed by Aristotle's statement that the observation of particulars presupposes a great deal of abstract theoretical knowledge. He thus endeavored to extract this theoretical knowledge from our experiences, a process that he titled "transcendental logic", and which he never actually describes or details. Although the bombastic name may somewhat intimidate us, we should note here in passing that there is nothing particularly transcendental about such an endeavor: the inference "x is impossible unless y is true; x is possible; hence y is true" is a rather basic case of *modus tollens*. Its premises are nothing more than empirical conjecture, as Salomon Maimon had observed.

And consider George Boole . He formulated the first extensional logical system, thus destroying by fiat the grand Aristotelian plan to logically establish empirical theoretical knowledge. Still, he tried to utilize his brilliant new logic to secure empirical science by probability and induction. However, in Boole it is already very clear that knowledge of the various probabilities used in inductive inference is an extra-logical, empirical matter, and hence that it is not logical reasoning alone that secures the foundations of science.

But it is Frege, as Agassi observes (Agassi 2018), who deserves credit as the first modern logician proper: he is so in virtue of the fact that he was the first to have abandoned altogether the Aristotelian program—securing empirical science—replacing it with the far more modest one of securing arithmetic by logical means alone. Then Russell discovered his paradox, and Gödel had sealed matters for this new program too, by demonstrating that, strictly speaking, it cannot be performed.

The bankruptcy of the two justificationist programs brings us back to the crucial question of the desired place of logic in science. What is its role there? What service does it provide to the scientist? Today it is hard to fathom, but until Popper arrived at the scene logic was solely a tool in the hands of dogmatists: a tool for establishing some truth (empirical or arithmetic), for justifying it. However, as Popper insisted, the only way to learn something about a universal empirical statement from a singular statement, is by observing that the latter refutes the former. Thus logic becomes refutation theory. As refutation theory logic finally returns to the Socratic role of methodology proper, methodology without guarantees for success, methodology that is free of its historical epistemological burdens, and indeed, as Bartley insisted, methodology that can one day be modified, at least in principle.

Consequently, in Popper, for the first time we find a view of science that is anti-foundationalist. It does not proceed from first principles (from essential definitions, or from immediate experiences), but rather from problems, that is from the challenge of explaining inconsistencies between our theories and our experiences. Indeed, as Popper had observed (echoing Aristotle) observation reports are theory-laden, and so (contra Aristotle) the theory which they presuppose, even when it is a priori, is conjectural too. This also freed logicians from the impossible burden of justifying induction: it is not merely that the task is now openly admitted as essentially hopeless, as Hume has already done, but it becomes essentially uninteresting, as science has no use for it.

Popper also famously used logic that is the theory of refutations, as a tool for demarcating science. This use was problematic since it sometimes gave the misleading impression that Popper intended to declare some theories (e.g. Newtonian physics) as more easily refutable than others (e.g. Adler's psychology) (Popper [1959] 2005). Clearly, refutability is not a property of theories. Rather critical mindedness is a property of speakers: the more critical minded we are towards a given theory, the more scientific it becomes. This way, logic is not a tool for demarcating science from pseudo-science, but rather an aid in distinguishing between the rational and the dogmatic.

References

Agassi, J.: Ludwig Wittgenstein's philosophical investigations. Springer, Cham (2018)

Aristotle: categories, on interpretation, prior analytics (Tredennick, H. trans.). Harvard University Press, Cambridge MA (1938)

Aristotle: posterior analytics, Topica (Forster, E.S. trans.). Harvard University Press, Cambridge MA (1960)

Bar-Am, N.: Extensionalism: the revolution in logic. Springer, Cham (2008)

Popper, K.: The logic of scientific discovery. Rouledge, London [1959] (2005)

Russell, B.: Mysticism and logic. Unwin Books, London [1910] (1963)

Nimrod Bar-Am is a philosopher of science that specializes in the history of logic and its relations with epistemology and metaphysics. He heads the Rhetoric and Philosophy of Communication Unit at the Communication Department of Sapir College, Israel. He published Extensionalism: the Revolution in Logic (Spinger 2008) and In Search of a Simple Introduction to Communication (Springer 2016). E-mail: ngbaram@gmail.com

Biology

Rehabilitation of Karl Popper's Ideas on Evolutionary Biology and the Nature of Biological Science

Denis Noble and Raymond Noble

1 Introduction

In this article we discuss what we consider to be significant contributions Karl Popper made to evolutionary biology and to the fundamental nature of biology as a science. We explain why his ideas were controversial and why they were forgotten and neglected until recently. The key argument is on how the agency of organisms mediate change, which we develop in five parts: (1) Karl Popper and Evolutionary Biology; (2) Can Biology be reduced to Chemistry? (3) How organisms can be agents in development and evolution; (4) Organisms as open systems; (5) Conclusions: are active and passive Darwinisms entirely separate?

2 Karl Popper and Evolutionary Biology

In 1986, Popper gave a lecture to The Royal Society in London in which he laid out his "New Interpretation of Darwinism." The history of that lecture is the subject of a recent book by Hans-Joachim Niemann (2014), and a relevant article by Eva Jablonka (2017).

D. Noble (✉)
Department of Physiology Anatomy & Genetics, University of Oxford, Oxford OX1 3PT, UK
e-mail: denis.noble@dpag.ox.ac.uk

R. Noble
Institute for Women's Health, University College London, London WC1E 6AU, UK

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_11

2.1 Passive and Active Darwinism

In his lecture, Popper distinguished between what he called “passive Darwinism” and “active Darwinism.” As these terms are not used today, they need an explanation, particularly because Popper referred to both of them as versions of “Darwinism” and even regarded his ideas as a refutation of Darwinism. However, in this respect Popper was not correct; it was a refutation of *neo*-darwinism, and not of the ideas proposed by Darwin. As we will show, Darwin was the first to develop both passive and active forms of his theories of evolution. In this, Darwin predated Popper.

Popper’s “passive Darwinism” is more or less identical with classical neo-Darwinism: the theory that random genetic variation and natural selection are entirely sufficient (*allmacht* in Weismann’s [1893] words) to explain evolution. “Passive” refers to the fact that the standard neo-Darwinist Modern Synthesis does not regard organisms as active agents in their evolution. Over the generations, they experience random variations in their genetic inheritance, but they are not viewed as *using* that randomness to generate functional variation that might be inherited. The organisms play no active part in evolution other than to pass these genetic variations on to the next generation. In this view, organisms are passive vehicles in gene transmission. In recent articles, we have outlined the opposing view that organisms themselves can, at least partially, guide evolution through the active choices that they make, and that this is achieved through the harnessing of stochasticity (Noble 2017; Noble and Noble 2017, 2018). The existence of such harnessing is an empirically testable process. The feedback control mechanisms determining the speed and location of genetic mutations and genome rearrangement are open to experimental investigation as dynamic biological processes. Furthermore, through such guidance (harnessing) the genetic variation is no longer random, but adaptive, even though randomness is used in the hypermutation and other cellular network processes involved. Those processes are also experimentally testable. Thus, evolution is an extension of the adaptive processes (physiology) by which organisms maintain integrity in response to environmental change.

There are therefore strong parallels between our work and that of Karl Popper. Popper wrote:

“I shall attempt to turn the tables completely on passive Darwinism… I shall claim that the *only* creative element in evolution is the activity of living organisms.” (Niemann 2014, 119). “Active Darwinism” is therefore equivalent to the theory that organisms have agency and make choices, which is the main theme of our recent papers. Those choices include selecting niches (niche selection theory) and which other organisms they interact with (social, including sexual, selection), and more recently, the discovery of aversion to cheating behaviour in populations of dogs (Essler et al. 2017) and monkeys (Brosnan and De Waal 2003).

Popper also regarded the “metaphor of ‘natural selection’” as “a theory of error elimination” (Niemann 2014, 120) rather than being creative of novelty itself. He saw it as a filter eliminating errors. On its own therefore, natural selection does not involve agency.

To understand this point, we should remember that Darwin contrasted natural selection with artificial selection, which is clearly dependent on choices made by organisms (the selective breeders). His 1859 book, *The Origin of Species*, begins with a chapter on "Variation under domestication". He noted that breeders of varieties of dogs, cats, fish, and plants were actively (consciously) choosing the characteristics of the varieties they had developed. They were doing so as *active* human (= artificial in this context) agents. In introducing the term 'natural selection' Darwin was, by metaphorical extension, attributing selection also to an essentially blind *passive* process, and specifically *in contrast to* active choice in organisms. This was his great achievement in his 1859 book *The Origin of Species*. His theory of natural selection, therefore closely corresponds to what Popper called passive Darwinism.

In his later book *The Descent of Man and Selection in Relation to Sex*, Darwin (1871) showed that what he attributed to active 'artificial' selection to humans also occurs as an evolutionary process in many animals, through the process of what he called sexual selection.[1] However, the co-discoverer of the concept of natural selection, Alfred Russel Wallace, disagreed with him, and subsumed sexual selection to natural selection. That was perhaps the first great philosophical error in the subsequent narrowing down of the theory of evolution by natural selection to eventually become the Modern Synthesis. It is an error that has continued to confuse modern biology to this day. Julian Huxley in his 1942 book *Evolution: The Modern Synthesis* did not even include sexual selection as a process driving evolution. Moreover, it is a common assumption by many supporters of the more dogmatic forms of the Modern Synthesis that organisms cannot be active agents. This aversion to agency explains why they go to great lengths to understand sexual selection purely in terms of natural selection. It also explains why they often also deny agency (active choice) to humans. We will deal with these issues in Sect. 3.

When Darwin realised that sexual selection is more like artificial selection, he realised that it is clearly an *activity* of organisms partially determining their evolution. Darwin recognized this difference as empirical since he wrote "with respect to animals very low in the scale, I shall have to give some additional facts under sexual selection, shewing that their mental powers are higher than might have been expected." (Darwin 1871, I, 35–36). Sexual selection is therefore a form of active Darwinism to use Popper's terminology. Specifically, Popper wrote "sexual selection is a refutation of natural selection." (Niemann 2014, 128) Darwin distinguished very clearly between the two and we believe Popper and Darwin would have agreed on this distinction.

Popper can therefore be seen to be following in the footsteps of the *original* position of Charles Darwin, in opposition to the neo-Darwinist view that everything in evolution is attributable to blind natural selection. Since Popper persisted in regarding his lecture as a "refutation of Darwinism" he may not have appreciated that Darwin had already made the distinction to which he was drawing attention, and that it was therefore neo-darwinism that was the real target of his lecture. Else, he recognised

---

[1] Actually, Darwin (1859) had already identified sexual selection briefly in *The Origin of Species*, 101–104.

that Darwin's position had been distorted by what had become generally regarded as 'Darwinism'. The Modern Synthesis is not the position developed by Darwin. It presents a gene-centric determinism, where physiological adaptation in the organism cannot be passed on through the germ line—genes make proteins and function, but proteins and function cannot alter the genes. The genome is viewed as a 'blueprint' for development and function, where change of that blueprint is dependent on the two ingredients of blind chance and natural selection. It established a kind of mechanistic dualism with a privileged role for the genome in function, and in doing so it removed agency from the organism. It locked the gene away in a box, free from the choices made by organisms.

2.2 Role of Indeterminacy

Popper saw that a complete determinism is incompatible with viewing organisms as agents making purposeful choices. Thus, he would have seen the significance of the role of harnessing stochasticity in creative responses to change and in making choices, which we have highlighted in our recent articles. However, he also recognised that indeterminacy, or blind chance, was not alone sufficient for an open, and thus creative system. It is not the unpredictability of events that creates agency. On the contrary, as we have argued, agency requires anticipation of change and the outcome of action. Yet, organisms harness stochasticity throughout biological function as the energy of creative change. Life constrains stochasticity, moulding it in function. It is the key ingredient of all physiology—from generating membrane potentials and synaptic function to releasing hormones, the beating of the heart, moving our limbs, and thinking.

In *The Open Universe: An Argument for Indeterminism*, Popper demonstrated that indeterminism is a necessary but not sufficient condition for emergence and openness (Niemann 2014, 70).

In the same exposition of Popper's ideas leading up to his Royal Society lecture, Niemann presents some other points that correspond well to the ideas of our work on agency. Summing up Popper, he concluded that "all life is problem solving. Acquiring new knowledge is always purposeful activity." (Niemann 2014, 90). Popper insisted that "in all cases the activity comes from outside of the DNA. The former 'centre of life' is rather a dead place." (Niemann 2014, 96). It is the cell that divides, not only the DNA (Niemann 2014, 98), and that it is "The cell… also managing the genome." (Niemann 2014, 101) This insight resembles that of Barbara McClintock, the discoverer of natural genetic engineering (Shapiro 2011) in saying that "the genome is an organ of the cell" (McClintock 1984).

Thus, Popper realised that the genome is the prisoner of the cell and organism and not the other way round. He also pointed out that "influences (on action) [are] traceable in hindsight… we are unpredictable but not irrational" (Niemann 2014, 110). Popper therefore arrived at many of the points we have made in recent articles.

In solving problems life can create solutions, else it cannot solve problems. The key problem is that of maintaining the integrity of life.

It would therefore be surprising if Popper had not also seen the obvious implication, which is that organisms harness stochasticity; otherwise, creative choice in behaviour would not be possible (Noble and Noble 2018). Thus, we are grateful to Hans-Joachim Niemann for directing us to Popper sources preceding his Royal Society lecture where he does clearly draw this conclusion. Some of the relevant texts occur in his dialogue with John Eccles *The Self and Its Brain* (Popper and Eccles 1977). Popper writes "New ideas [*in statu nascendi*] have a striking similarity to genetic mutations" and continues "describing 'the process with respect to new ideas and to free will decisions'(Popper and Eccles 1977, 540) as randomly produced proposals followed by selection based on standards coming from the world" (Niemann 2012, S510–S546). Popper arriving at this conclusion is a logical outcome of his earlier (1973) conclusion that "indeterminism is not enough" (Popper 1945, vol. 2, 210, 1973).

Certainly, we agree that indeterminism is not enough for creative or purposeful agency. In standard evolutionary biology, following the Modern Synthesis, stochasticity generates random genetic variations. But in the standard theory this is not directed in a functional way. These chance variations may or may not confer any advantage on the organism, although those that do will more likely be retained in the 'gene pool'. In contrast, the direction of agency comes because organisms *harness* stochasticity in functional ways. Thus, the immune system creates hypermutation in highly targeted regions of the DNA sequence of immunoglobulin proteins. Under stress, bacteria also use hypermutation to resist antibiotics and to counter other forms of genetic loss.

Nevertheless, while Popper envisaged "the cell… also managing the genome," (Niemann 2014, 101), he does not seem to have arrived at the details of the comparison with hypermutation in the immune system. This is not surprising since the discovery of some of the detailed molecular mechanisms of somatic hypermutation occurred in 1999 after his death in 1994 (Muramatsu et al. 1999; Li et al. 2014). There may also have been a puzzle regarding the molecular mechanism of hypermutation. Increasing the natural mutation rate by a factor of up to 10^6 (a million-fold increase!) must have seemed implausible. But this is no longer so astonishing since it is also roughly the order of magnitude difference between the natural mutation rate in DNA copying before and after repair by cellular editing mechanisms (Noble 2018). Mismatch DNA repair is indeed suppressed during somatic hypermutation. Recent research has therefore shown that there is no difficulty in accounting for hypermutation rates of up to a million times normal. All that is required is to inhibit the error-correcting process in the relevant part of the genome to bring about targeted change. Thus, the stochasticity can be released from constraint in a targeted way. We have argued that it is this targeted process of mutation that gives direction, or agency, to the organism in response to environmental change.

3 Can Biology Be Reduced to Chemistry?

Why was Popper's Medawar lecture never published by The Royal Society journals? A possible explanation is that, following the lecture, Popper engaged in extended discussion with Max Perutz on the question whether Biology could be reduced to Chemistry (Niemann 2014, 62–66). On this question Popper and Perutz were in complete disagreement.

Perutz, a molecular biologist who shared the 1962 Nobel Prize for Chemistry with John Kendrew for their studies of the structures of haemoglobin and myoglobin, presented the reductionist case that had gained hold in biological science. Popper disagreed, as one would expect from the earlier quote:

> in all cases the activity comes from outside of the DNA. The former 'centre of life' is rather a dead place.

Popper's statement is surely correct. DNA does nothing outside a cell. A cell is required for both copying and error correction[2] and is therefore the minimal form of a living system. In order for proteins to be made, copying of the relevant DNA sequences needs to be activated by the cellular processes. A cell is also required to perform the extensive error-correction necessary to ensure faithful replication of DNA to pass on to the next generation. Else random mutation would rapidly destroy its integrity.

Perutz clearly never understood this point since, after Popper's death, he still claimed in a paper entitled *Darwin was right* that "DNA is the score of the music played by the cell" (Niemann 2014, 66). However, the real question is not merely who plays the score, but who or what writes and maintains it. The tune played by the cell is not produced by the DNA, but by the processes of the cell that use the DNA to create the music of life. In this sense, the DNA is not the score; it is an instrument used by the cell.

Thus, the Central Dogma of Molecular Biology, from which Perutz like many others argued, is a viewpoint, and arguably a mistaken one. It is simply a chemical fact about DNA acting as templates for amino acid sequences. It does not alone have the capacity for genetic causation, nor does it exclude macro-level control feeding back onto the genome (Noble 2018). Indeed, such feedback is essential for living cells to work. Once it is accepted that such feedback exists, it opens the potential for agency in evolution at the cellular level.

Research on this issue has greatly progressed since Popper's time. We now understand much better the ways in which the genome is controlled by higher levels of organisation. These controls can be represented mathematically since they determine the initial and boundary conditions for any molecular level representation of biological processes. A concise way of stating this fact is what we have called the principle of biological relativity, which is the statement that, a priori, there is no privileged level of causation (Noble 2012). An important question then is what characterises the physiological properties of the boundaries between the levels of organisation.

---

[2] There are many demonstrations of this fact. For a brief account see Noble (2018).

We have attempted to clarify that question with a variety of examples of experimental and modelling work on the boundaries between different levels Noble et al. 2019). The answer is that there are several different ways in which causation can be transmitted through the boundaries. Nature seems to have been opportunistic in employing whatever mechanisms were open for evolution to exploit.

The principle of biological relativity (Noble 2012) was derived from the general principle of relativity used in physics, and was inspired by a meeting on Top-Down causation organised by the mathematician and cosmologist, George Ellis. The principle has now been extended back, by Ellis, into a form of relativity of causation in physics itself (Ellis 2020):

> no level is a fundamental level with priority over the others, and particularly there is not a primary one at the bottom level. This is just as well, because there is no well-established bottom-most physical level to which physics can be reduced. Every emergent level equally represents an effective theory. (Murugan et al. 2012)

This means that even chemistry cannot be reduced to physics and, even more so, biology cannot be reduced to chemistry. The fundamental reason for these conclusions have also recently been explored by Stuart Kauffman:

> the becoming of biospheres falls entirely outside the Newtonian Paradigm. The reason, as we shall see, is that the very phase space of biological evolution – which includes biological functions – persistently evolves in ways that we cannot even prestate, let alone predict. Without a prestated phase space, we can write no law of motion in the form of differential equations, hence we cannot integrate the equations we do not have. Thus, no laws at all entail the stunning unfolding of our, or any, biosphere in the universe. (Kaufmann 2020)

The arguments in the papers of Ellis and Kauffman can also be viewed as further exploration of the consequences of the existence of open systems, to which we will return in Sect. 4.

4 How Organisms Can Be Agents in Development and Evolution

Popper's argument with Perutz naturally leads to a related question: can mental processes be causal? Clearly, if conscious intention is a mere illusion with no causal power then organisms cannot be active agents in their development and evolution. Popper's and Darwin's arguments for social, including sexual, selection would then carry no weight. That is precisely what many neo-darwinists and strong reductionists believe, which is why we suspect that this issue was at the heart of the Popper-Perutz argument.[3]

The demonstration that there are several ways in which causation can be transmitted between boundaries provides an answer to the commonly held view that

---

[3] A note for future historians: this question might be settled when the Popper archive at The Hoover Institution, Stanford, becomes open to researchers in 2029.

mental events cannot be causal. We will show that, as this view is expressed by Jaegwon Kim (perhaps the strongest advocate of the non-causal view), it is incompatible with multi-scale causation in open, such as living, systems. Far from it being the case "that physical causes exclude mental states from causally contributing to the behavior", even a rigorous mathematical analysis of physical causation, e.g. in differential equation models, is necessarily incomplete without the contextual, including mental, processes. The forms of causation involved are complementary, they necessarily mesh with each other (Noble and Noble 2020).

We will illustrate this point by considering some of Jaegwon Kim's central arguments and how the complementarity of the forms of causation, together with the harnessing of stochasticity, can be shown to deal with those arguments.

Kim poses the dilemma very clearly:

> The problem of determinism threatens human agency, and the challenge of scepticism threatens human knowledge. The stakes seem even higher with the problem of mental causation, for this problem threatens to take away both agency and cognition. (Kim 2000, 32)

One of the most powerful reasons Kim adduces to justify this threat is the problem of causal exclusion:

> Suppose then that mental event *m*, occurring at time *t*, causes physical event *p*, and let us suppose that this causal relation holds in virtue of the fact that m is an event of mental kind *M* and *p* an event of physical kind *P*. Does *p* also have a physical cause at *t*, an event of some physical kind *N*? To acknowledge mental event *m* (occurring at *t*) as a cause of physical event *p* but deny that *p* has a physical cause at *t* would be a clear violation of the physical closure of the physical domain………the physical cause therefore threatens to exclude, and pre-empt, the mental cause…..The antireductive physicalist who wants to remain a mental realist, therefore, must give an account of how the mental cause and the physical cause of one and the same event are related to each other. (Kim 2000, 37)

Our answer will be in three parts:

I. In organisms there can be no causal closure at the micro-level.
II. The absence of causal closure depends on the relation of causes between and within levels of organisation.
III. The existence of stochasticity and its harnessing enable multiple scenarios to be anticipated by organisms from which ones that most closely instantiate the possible reasons for an action can be selected.

4.1 No Causal Closure

Organisms are composed of levels of organisation nesting within each other, and with the environment. For there to be even the possibility of causal closure we must first define the boundary of the system we wish to investigate in any causal model. At whatever level we do that, whether molecular, cellular, tissue, organs, systems or the organism as a whole, all levels within the chosen boundary will display dynamic processes since organisms are not static. They are never at equilibrium. We cannot therefore represent those processes by static algebraic equations. We note this point because many models used in evolutionary biology and in social sciences such as economics do use static algebraic equations. The problems of causal closure and the openness of the system are then hidden (Noble and Noble 2020).

Most often in non-static models the dynamics are represented by differential equations or by their equivalent. In all cases the equations by themselves do not have solutions until we add the initial and boundary conditions. Those come from the history and the organisation at all levels with which interactions can occur, including the social level at which interpersonal relationships and ideas matter. Those initial and boundary conditions cannot be restricted to events and organisation within the level we have chosen to represent. In open systems, there cannot be causal closure within any level alone. These mathematical considerations lie at the heart of the formulation of the principle of biological relativity.

4.2 The Relation of Causes Between and Within Levels of Organisation

An important outcome of multilevel representations in biology is that the forms of causation up and down across the boundaries between the levels are not equivalent. The history and organisation at higher levels do not determine or contradict the dynamic equations used to represent activity in lower-level processes, they constrain those dynamics, in much the same way that the form and elasticity of the boundaries containing a gas constrain the movements of individual gas particles to produce the overall parameters of pressure, volume and temperature. Someone focussed on the equations for the lower-level dynamics may well not realise that and think that those equations contain all that may be required for causal closure. But they can only do that by not being aware that they will have imported factors (usually in the form of initial and boundary conditions) that are not themselves dynamic in the way in which the dynamic equations are. We consider that this is what is missing from Kim's accounts, and why causal closure is impossible. The forms of causation do not compete, they are necessarily complementary; they mesh together. To modify the relevant part of Kim's statement: the physical cause cannot therefore threaten to exclude, and pre-empt, the mental cause since the physical form of causation is itself necessarily constrained by higher level causation, which is not itself another form of dynamic causation.

4.3 The Existence of Stochasticity and Its Harnessing

In recent articles we have shown that organisms do not just experience stochasticity. They use it to explore options and select from those options (Noble 2017; Noble and Noble 2017, 2018, 2020). This is one of the essential bases of agency, as Popper also saw (Sect. 1.2 above). The reason is that it enables organisms to explore and anticipate many options. To answer Kim again, there isn't just one series of *p*'s there can be an indefinite number. It is by selecting from amongst them that organisms can discover solutions to problems posed by the environment and in interaction with other organisms. They can select the ones that most closely instantiate the possible reasons for an action. Those possible reasons form the cultural context within which organisms with agency can act.

5 Organisms as Open Systems

In his 1945 book, *The Open Society and Its Enemies*, Karl Popper contrasted closed, deterministic views of society with open and creative system views. In the open-system perspective, ideas are a vital ingredient of the dynamics of change, involving creative agency; in the closed view, the principal causes of transformation are embedded in the system, determinedly driving it forward. These opposing perspectives have profound consequences. At its extreme, the closed view makes us prisoners of our determinate existence—at best, we can only mitigate the outcome, or "lessen the birth-pangs" by understanding the nature of that change. In the open view, we can be arbiters of our destiny, using our understanding to bring about creative evolution.

The question is not whether we are free but the extent to which we are open and how we can be free agents in our destiny. Where Marx concluded "It is not the consciousness of men that determines their existence, but their social existence that determines their consciousness." he sees a mostly unidirectional causal chain or at least a profoundly weighted one. Yet, consciousness is a vital capacity of our material existence; it is an essential part of our social reality and not an illusory epiphenomenon. It is not merely a product of our social presence. It engages in our choices, individually and socially. We are not free from our material existence, nor do we exist apart from it; clearly, we depend upon it; but we have agency within it. As Popper puts it in countering the hard historicist Marxist view "the future depends on ourselves, and we do not depend on any historical necessity." (Popper 1945, I, 3).

Thus, for determinists, or historicist accounts, forces beyond our control overwhelmingly govern our destiny. It is as if we can see nature unfold, yet have no way to use that vision to alter its direction. In this view, consciousness is a product and not a player in our history. Yet, agency and awareness play vital roles in the integrity of living organisms. We are players in our history and not merely products of it.

The parallels with biology are not surprising; being social is part of our biology. A closed system has no creativity, and this was Descartes' problem. If organisms are viewed as mere machines, working like clocks, then where is the agency? If we are automata, then where is our will? We end with a body-soul, or body-mind split, and this, in turn, creates the problem of how one (immaterial) can be influenced by the other (material).

We have argued that treating organisms as closed systems leads to another, but similar kind of dualism—a materialist dualism, or something within the system that drives the system. This dualism unfolds as genetic determinism, with genes as driving agents. We become prisoners of our genes. A bit of the system that is so important that organisms become mere 'vehicles' in its transgenerational transmission, or as Dawkins put it in *The Selfish Gene*:

> We are survival machines – robot **vehicles** blindly programmed to preserve the selfish molecules known as genes. This is a truth which still fills me with astonishment. (Dawkins 1976)

This 'truth' separates the gene (the driver) from its vehicle (the organism), thus creating an unnecessary materialist dualism. It fills us with astonishment that it is presented as 'truth'.

Popper's book was a robust defence of the open society. It is not therefore surprising that he should have championed the role of agency in biology. All life depends on the exchange of matter and energy with the environment, including social arrangements with other organisms. We can represent this openness as a matter of degree, dependent on increasing complexity of self-organisation, represented diagrammatically in Fig. 1.

The principle behind this diagram is that the sensitivity to variation dependent on the environment of each level of organisation increases as we move up from the molecular level through to the social level. At the bottom, DNA, like all molecules in isolation, cannot be said to possess agency. We only begin to encounter agency, in the sense of self-maintenance of an organised system, when we move to levels higher than the molecular.

Notice that we have specified the karyotype as a level of organisation. We have done that to recognise one of the most important discoveries in genome wide sequence

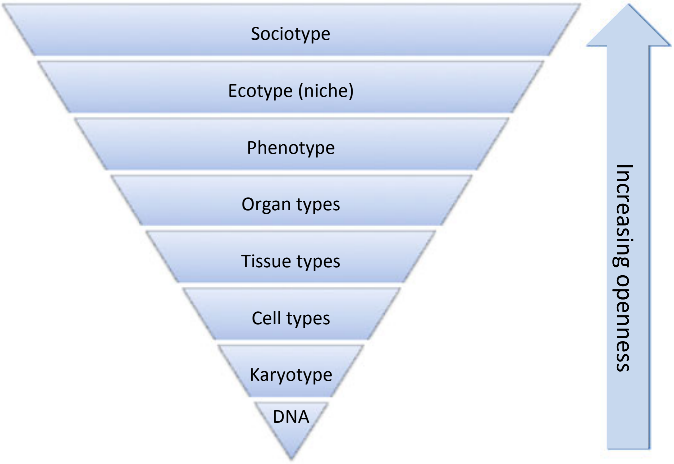


Fig. 1 Levels of organisation in organisms. The width of the cone represents the degree of openness. At every level, the system is open, yet functionally constrained by the levels above, which are more variable and malleable

studies, which is that the associations of individual genes (DNA sequences) with phenotypes, healthy or diseased, is generally very low. The hope of gene-centric biology delivering the basis of cures for common diseases has not materialised. We now know that it is the complete genome that is important to the overall genetic contribution to inheritance: the omnigenic hypothesis (Boyle et al. 2017). In his book *Genome Chaos* Heng (2019) goes even further and calls for an integrated view of the karyotype, which refers to the complete chromosomic structure as itself an organised system. The idea is not just chromosomes as a set of individual genes, but rather the activity of chromosomes in genome control and rearrangement. The karyotype is therefore a distinct level of organisation, arranged as a 4D structure rather than a 1D sequence. The karyotype is itself sensitive to control from higher levels.

The other levels are well-known already from many similar diagrams of the hierarchical organisation of organisms. But our top level, sociotype, does require some comment. The social interactions are where conscious agency is generated. The fluidity and contingency of the sociotype is also an important source of stochasticity forming the clay out of which functional novelty can be generated.

Another way of illustrating the degrees of openness of biological systems is to represent the evolution of openness as a time series of major transitions. We attempt to do this in Fig. 2.

The important point to note about this figure is that each transition depends on the evolution of the prior transitions. There is a ratchet effect (Hoffmann 2012). Each time a major transition is achieved, it opens the way for the next transition. The nature of the evolutionary process also changes. Evolution itself evolves (Noble et al. 2014).

This process of emerging evolution is an old idea in evolutionary biology, explored notably in Maynard Smith and Szathmary's (1995) book *The Major Transitions in Evolution* and more recently Ginsburg and Jablonka's (2019) book *The Evolution of*

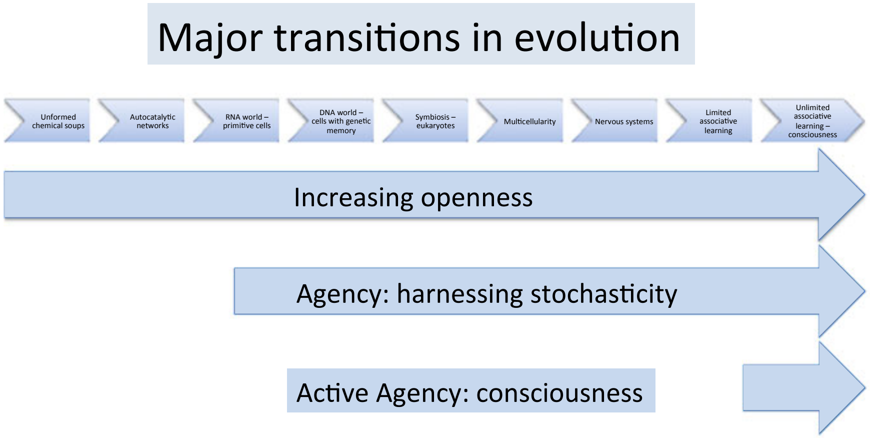


Fig. 2 Evolution of organisms represented as a possible time sequence with major transitions

the Sensitive Soul. It also goes back at least to Lamarck's idea of *Le pouvoir de la vie* (the force of life—Lamarck 1809). Lamarck has been widely ridiculed for his ideas and this one is no exception. However, Lamarck was not a vitalist. He was firmly a materialist (Pichot 1994). The current idea of a ratchet process in the development of major transitions corresponds well to what he had in mind, which was that each stage in the evolution of organisms created a further way up the ladder of complexity. It is widely thought that Lamarck also thought that the ladder concept best represented the transformation of species with no branching. This is not correct. Lamarck himself changed his mind and replaced his ladder with a tree of life (Noble 2020).

The specific point about such a diagram that is relevant to our article is that the transitions all mark the increasing development of openness. This is particularly true for the later stages, where we have followed Ginsburg and Jablonka in distinguishing as separate stages the development of nervous systems, on the basis of which associative learning can develop, later to become unlimited associative learning, which is their proposed marker for the development of consciousness: the last transition in Fig. 2.

6 Conclusions: Are Active and Passive Darwinisms Entirely Separate?

So far in this article we have expounded and interpreted Popper and Darwin on the issue of the active-passive distinction of Popper and the artificial-natural distinction of Darwin. We have shown that they correspond. We have shown that Popper's active Darwinism is Darwin's artificial selection and Popper's passive Darwinism is Darwin's natural selection. The correspondence is very close.

But we now have to row back somewhat. Like many distinctions in language and philosophy they are not so clear cut as may first appear. What, after all, is natural selection? Darwin saw the contrast as between what humans do artificially through selective breeding and what happens naturally when the environment acts as the passive filter of natural selection.

But what creates that environment?

Consider this:

Tools and language facilitate agency. We use machines and communication for reasons—to do something or to express something. Organisms use the first to do things and the second to communicate. With tools, organisms obtain food and build protection from the physical environment. Language and writing improve communication and enable ideas to be explored, transmitted and transformed across generations and between groups or individuals. Using language enables communication and understanding of intention. With tools and language, humans created civilisations and extended abstract thought through literature and art. This creativity, in turn, influences the way we perceive the world. The built environment and the psychological texture of society is the explorative embodiment of niche creation and through which selective pressure affects human evolution. We are not merely hunter-gatherers in a concrete jungle; we are evolving organisms in a created niche. What we expect of each other and ourselves affects our physiology. Sometimes we suffer as a result. If natural selection is the measure of fitness to survive and reproduce, then we must ask what it is that is doing the selecting, and what it is that is being selected. If it is the environment, then clearly, we also consciously create that environment. Humans are not alone in doing this.

Just as we have selected dogs, cats and other domesticated and farm animals, so we also choose each other, as partners, as friends, and often with whom we work. Our social being shapes us as individuals, just as we form our social being. The relationship between consciousness and nature is intertwined and not separate.

All the time, over different time scales the environment itself evolves as a consequence of the agency of organisms. We illustrate this in Fig. 3. Popper understood and emphasised the two-way process:

> Our minds are the creators of world 3; but world 3 in its turn not only informs our minds, but largely creates them. The very idea of a self depends on world 3 theories, especially upon a

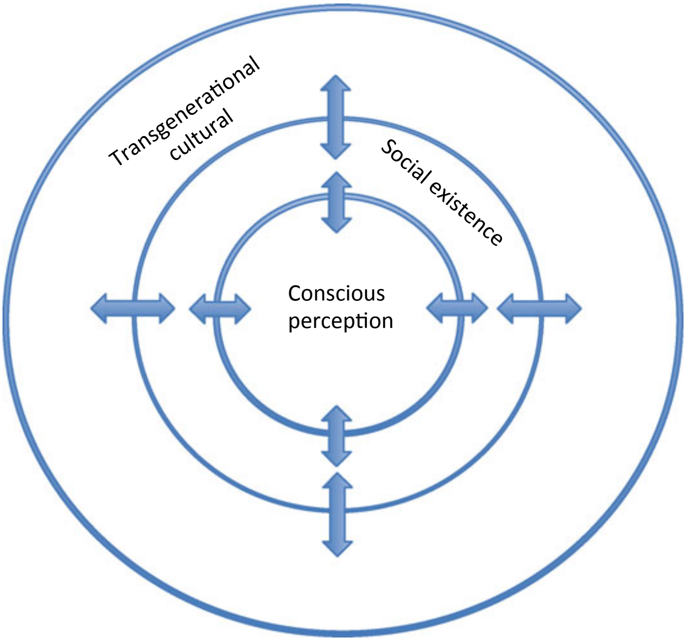


Fig. 3 Diagram of functional interactions between social existence and conscious organisms with agency. Agents with conscious perception interact across functional boundaries with their social existence, which in turn facilitates interacts through the transgenerational cultural inheritance, allowing the creation of ideas, viewpoints, opinions, attitudes and actions. We have used double-headed arrows to emphasise that there is no privileged circle of interaction. There is a continuous two-way interaction. The transgenerational cultural inheritance forms what Popper (1978) referred to as World 3

> theory of time which underlies the identity of the self, the self of yesterday, of today, and of tomorrow. The learning of a language, which is a world 3 object, is itself partly a creative act and partly a feedback effect; and the full consciousness of self is anchored in our human language. (Popper 1978)

The environment is what we, the organisms, have created, including the transgenerational cultural inheritance. Organisms have also completely changed the physical and chemical environment. The environment of the earth today is nothing like the environment of the earth when life first evolved. So, over all the more than 3 billion years since then we, organisms, have altered the environment in almost all relevant respects. Even natural selection eventually operates as it does today through the actions of organisms. The world of nature (natural) is not merely physical, it is biologically functional (selective).

Of course, there is a difference of time scale. At any one period of time we can distinguish between natural and artificial selection. Our concluding point does not invalidate the distinctions both Popper and Darwin were drawing. But there is nevertheless an important process to add to what we have described in this article. There is a continuous circular interaction between the activity of organisms and the development of the environment. That, in turn, becomes the basis of natural selection. Even natural selection is therefore not entirely passive. Humanity itself is now the greatest driver of evolution through rapid alteration of the global environment (Corning 2020).

Acknowledgements No funding supported this work. We have no conflicts of interest.

References

Boyle, E.A., Li, Y., Pritchard, J.K.: An expanded view of complex traits: from polygenic to omnigenic. Cell **169**, 1177–1186 (2017)

Brosnan, S.F., De Waal, F.B.: Monkeys reject unequal pay. Nature **425**, 297–299 (2003)

Corning, P.: Beyond the modern synthesis: a framework for a more inclusive biological synthesis. Prog. Biophys. Mol. Biol. **153**, 5–12 (2020)

Darwin, C.: The origin of species. (1859) (Macmillan Collectors Library edition (2004))

Darwin, C.: The descent of man, and selection in relation to sex. Murray, London (1871) I: 35–36

Dawkins, R.: The selfish gene. Oxford University Press, Oxford (1976)

Ellis, G.F.R.: The causal closure of physics in real world contexts. Nous (2020) https://aeon.co/essays/heres-why-so-many-physicists-are-wrong-about-free-will

Essler, J.L., Marshall-Pescini, S., Range, F.: Domestication does not explain the presence of inequity aversion in dogs. Curr. Biol. **27**, 1–5 (2017)

Ginsburg, S., Jablonka, E.: The evolution of the sensitive soul: learning and the origins of consciousness. MIT Press, Cambridge, MA (2019)

Heng, H.: Genome Chaos, rethinking genetics, evolution and molecular medicine. Academic Press, London (2019)

Hoffmann, P.M.: Life's ratchet: how molecular machines extract order from chaos. Basic Books (2012)

Huxley, J.: Evolution: the modern synthesis. London, George Allen and Unwin (1942)

Jablonka, E.: The evolutionary implications of epigenetic inheritance. Interface Focus. 7, 20160135. http://dx.doi.org/10.1098/rsfs.2016.0135 (2017)

Kauffman, S.: Eros and logos. J. Theor. Hum.Ities **25**(3) (2020)

Kim, J.: Mind in a physical world: an essay on the mind-body problem and mental causation. MIT Press, Cambridge, MA (2000)

Lamarck, J-B.: Philosophie Zoologique, original edition of 1809 with introduction by André Pichot. Flammarion, Paris (1809, 1994)

Li, Z., Woo, C.J., Iglesias-Ussel, M.D., Ronai, D., Scharff, M.D.: The generation of antibody diversity through somatic hypermutation and class switch recombination. Genes Dev. **18**, 1–11 (2014)

Maynard, Smith J., Szathmary, E.: The major transitions in evolution. Oxford University Press, Oxford (1995)

McClintock, B.: The significance of responses of the genome to challenge. Science **226**, 792–801 (1984)

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, K., K, Davidson, N.O. et al.: Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J. Biol. Chem. **274**, 18470–18476. (1999)

Murugan, J., Weltman, A., Ellis, G.F.R. (eds.): Foundations of space and time: reflections on quantum gravity. Cambridge University Press, Cambridge (2012)

Niemann, H-J.: Nachwort des Herausgebers. In: Wissen und das Leib-Seele-Problem, edited by K. R. Popper (Mohr Siebeck, Tubingen), pp. S510–S546; (afterword of the publisher; K.R. Popper, Knowledge and the mind-body problem). Sections 31–33 (2012)

Niemann, H.-J.: Karl Popper and the two new secrets of life. Mohr Siebeck, Tübingen (2014)

Noble, D.: A theory of biological relativity: no privileged level of causation. Interface Focus. **2**, 55–64 (2012)

Noble, D.: Evolution viewed from physics, physiology and medicine. Interface Focus. **7**, 20160159 (2017)

Noble, D.: Central dogma or central debate? Physiol. **33**, 246–249 (2018)

Noble, D.: Charles Darwin, Jean-Baptiste Lamarck, and 21st century arguments on the fundamentals of biology. Progress in Biophysics and Molecular Biology. 153, 1–4. https://doi.org/10.1016/j.pbiomolbio.2020.02.005 (2020)

Noble, D., Jablonka, E., Joyner, M.J., Muller, G.B., Omholt, S.W.: Evolution evolves: physiology returns to centre stage. J. Physiol. **592**, 2237–2244 (2014)

Noble, R., Noble, D.: Was the watchmaker blind? Or was she one-eyed? Biology **6**(4), 47 (2017)

Noble, R., Noble, D.: Harnessing stochasticity: how do organisms make choices? Chaos **28**, 106309. https://doi.org/10.1063/1.5039668 (2018)

Noble, R., Noble, D.: Can reasons and values influence action: how might intentional agency work physiologically? Under Rev. J. Gen. Philos. Sci. (2020)

Noble, R., Tasaki, K., Noble, P.J., Noble, D.: Biological Relativity requires circular causality but not symmetry of causation: so, where, what and when are the boundaries? Frontiers in Physiology **10**, 827 (2019)

Pichot, A.: Introduction, in Philosophie Zoologique. Flammarion, Paris (1994)

Popper, K.R.: The open society and its enemies. Routledge, London (1945)

Popper, K.R.: Indeterminism is not enough. Encounter **40**(4), 20–26 (1973)

Popper, K.R.: Three worlds. University of Michigan Tanner Lecture. https://tannerlectures.utah.edu/_documents/a-to-z/p/popper80.pdf. (1978)

Popper, K.R., Eccles, J.C.: The self and its brain. New York, Springer International (1977)

Saribasak, H., Gearhart, P.: Does DNA repair occur during somatic hypermutation? Seminar. Immunol. **24**, 287–292 (2012)

Shapiro, J.A.: Evolution: a view from the 21st century. Pearson Education Inc, Upper Saddle River, NJ (2011)

Weismann, A.: Die Allmacht der Naturzüchtung; eine Erwiderung an Herbert Spencer. Jena, Fischer, (the omnipotence of natural breeding; a reply to Herbert Spencer) (1893)

Denis Noble is Emeritus Professor at the University of Oxford, UK. In 1960 he was one of the first biologists to use mainframe computers to model biological organs. He has been involved in the development of mathematical modelling of biological systems ever since those early days. His most recent books are *The Music of Life* (OUP 2006) and *Dance to the Tune of Life* (CUP 2016) in which he explains why standard theories of evolution do not account for active agency in organisms. E-mail: nobleoxford@gmail.com

Raymond Noble is an Honorary Senior Lecturer in the Institute for Women's Health at University College London, UK. He trained as a zoologist and neuroscientist working on sensory processing by neural networks. He is a biologist and medical ethicist studying the nature of causation in biological systems and in how organisms make creative choices. He was former Deputy Dean of Life Sciences and Graduate Tutor in Women's Health at UCL.

Agency in Evolutionary Biology

Philip Madgwick

1 Why Does Evolutionary Biology Give a Privileged Role for DNA?

Contemporary evolutionary biology is a vast and loosely-connected discipline, so it is very hard to give an all-encompassing account of what it is all 'for', but I am nevertheless going to try. The vast majority of evolutionary biologists are working within a tradition that stems from Charles Darwin's (1859) *The Origin of Species* (hence the label 'Darwinism').[1] However, the contemporary tradition of evolutionary biology has been arrived at after a great discontinuity, which has been described as 'the eclipse of Darwinism'.[2] In this way, the contemporary tradition is often considered to have its foundations laid after the discontinuity in the research tradition stemming from Julian Huxley's (1942) *The Modern Synthesis*. To many critics of contemporary evolutionary biology—not least those like Noble (2006, 2016) concerned with the privileged role of DNA in evolutionary theory, the 'Neo-Darwinism' in *The Modern Synthesis* was where it all went wrong. I am going to address what happened at this critical juncture circuitously, by following the chain of reasoning from the statement of a problem that evolutionary theory sets out to explain and to the capitulation of 'the privileged role of DNA' on the way to its resolution.

P. Madgwick (✉)
Milner Centre for Evolution Department of Biology & Biochemistry, University of Bath, Bath, UK
e-mail: philipgmadgwick@gmail.com

[1] "*The point I want to make now is that all attempts to answer that question* [evolution, especially of humans] *before 1859 are worthless and that we will be better off if we ignore them completely*" (Simpson 1966, p. 1).

[2] This description is a chapter heading of Huxley's (1942) *The Modern Synthesis.*

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_12

1.1 What Is Evolutionary Biology All About?

Although Darwin's *The Origin of Species* represents the intellectual birth of mainstream evolutionary biology, the question at the heart of this book was a much older one: the problem of adaptation.[3] Yet, the way of approaching the problem was comparatively contemporary. Following in the tradition of British empiricism,[4] Darwin sought to explain adaptations with reference to features that were 'external' to the organism. The choice of this approach was heavily influenced by William Paley, who was by no means the originator of this externalism, but was amongst its most effective and influential advocates. Paley contended that the reason why the forms and behaviours we see in nature are one way rather than another has nothing to do with the individuals in question.[5] Instead, Paley argued, the forms and behaviours must be explained by the existence of a Creator. Darwin naturalised Paley's teleological argument and repurposed it to support his theory of adaptation through evolution by natural selection, which is famously based on a great analogy to animal and plant breeding.[6] *The Origin of Species* is, for the most part—and almost to the point of

---

[3] There are numerous books that frame the question of adaptation within a pre-Darwinian historical context, which trace adaptationist thinking back to the Ancient Greeks (Bowler 1983; Riskin 2016; Stott 2012), especially Aristotle who is even accredited by Darwin (1859) in later editions of *The Origin of Species*.

[4] I am using this term to refer particularly to the philosophical perspective epitomised by David Hume. These views differ from what John Locke referred to as 'continental rationalism', which is equally epitomised by Rene Descartes. 'British empiricism' is a contemporary term, which I use following (Godfrey-Smith 1996).

[5] "*If an account must be given of the contrivance which we observe; if it be demanded, whence arose either the contrivance by which the young animal is produced, or the contrivance manifested in the young animal itself, it is not from the reason of the parent that any such account can be drawn. He is the cause of his offspring in the same sense as that in which a gardener is the cause of the tulip which grows upon his parterre, and in no other. We admire the flower; we examine the plant; we perceive the conduciveness of many of its parts to their end and office; we observe a provision for its nourishment, growth, protection, and fecundity; but we never think of the gardener in all this. We attribute nothing of this to his agency; yet it may still be true, that without the gardener we should not have had the tulip: just so it is with the succession of animals even of the highest order. For the contrivance discovered in the structure of the thing produced, we want a contriver. The parent is not that contriver. His consciousness decides that question. He is in total ignorance why that which is produced took its present form rather than any other. It is for him only to be astonished by the effect*" (Paley 1802, p. 34).

[6] "*One of the most remarkable features in our domesticated races is that we see in them adaptation, not indeed to the animal's or plant's own good, but to man's use or fancy. … we must, I think, look further than to mere variability. We can not suppose that all the breeds were suddenly produced as perfect and as useful as we now see them; indeed, in many cases, we know that this has not been their history. The key is man's power of accumulative selection: nature gives successive variations; man adds them up in certain directions useful to him. In this sense he may be said to have made for himself useful breeds. The great power of this principle of selection is not hypothetical. It is certain that several of our eminent breeders have, even within a single lifetime, modified to a large extent their breeds of cattle and sheep. In order fully to realise what they have done it is almost necessary to read several of the many treatise devoted to this subject, and to inspect the animals*" (Darwin 1859, p. 23–24).

tediousness, a great catalogue of evidence collected from 'the many treatise devoted to this subject' of selection by breeders. But, in all important senses, Darwin left unchanged the externalist style of reasoning that was championed by Paley wherein individuals have adaptations because of some feature external to those individuals. Such externalism was in great contrast to the thrust of pre-Darwinian evolutionary theory, especially stemming from Lamarck (and continental rationalism).[7] So, what is the aim of evolutionary biology? In sweeping terms: to explain adaptation. But, to the extent that a research tradition is both a problem and a way of approaching that problem,[8] the problem of evolutionary biology is also set within an externalist method of enquiry.

As I have eluded to already, Darwin did not instigate a successful research tradition within his own lifetime—though he was nonetheless well-respected.[9] Instead, there were many apparently insurmountable criticisms, though it is interesting that the constructed history of this time by evolutionary biologists tends to focus on one: "*The biggest blank on the evolutionary map, however, concerned variation and its inheritance. The theory of mutation on a mendelian basis is the first adequate attempt to fill the gap*" (Huxley 1942, p. 109). Or, as was remarked after *The Modern Synthesis* (from which the quote above is taken) was published: "*The question Darwin failed to answer was actually a simple one. Survival of the fittest* what*?*" (Alexander 1979, p. 23). Progress toward an answer to this question started with the studies of the mechanism of inheritance by Gregor Mendel, which led to the development of population genetics. For those involved in this new field, the world of individual organisms rapidly becomes reconceptualised in terms of a genetic accounting.[10] Underlying this

---

[7] "*All major evolutionary theories before Darwin, and nearly all important versions that followed his enunciation of natural selection as well, retained fealty to an ancient Western tradition, dating to Plato and other classical authors, by presenting a fundamentally "internalist" account, based upon intrinsic and predictable patterns set by the nature of living systems, for development or "unfolding" through time*" (Gould 2002, p. 160).

[8] "*As the student proceeds from his freshman course to and through his doctoral dissertation, the problems assigned to him become more complex and less completely precedented. But they continue to be closely modelled on previous achievements as are the problems that normally occupy him during his subsequent independent scientific career. One is at liberty to suppose that some—where along the way the scientist has intuitively abstracted rules of the game for himself, but there is little reason to believe it. Though many scientists talk easily and well about the particular individual hypotheses that underlie a concrete piece of current research, they are little better than laymen at characterizing the established bases of their field, its legitimate problems and methods. If they have learned such abstractions at all, they show it mainly through their ability to do successful research. That ability can, however, be understood without recourse to hypothetical rules of the game*" (Kuhn 1962, p. 47).

[9] This respect can be seen in Darwin's 'major funeral' in Westminster Abbey, where he was honoured more as a public intellectual than as the father of evolutionary biology (Bowler 1984; Desmond and Moore 1991; Gould 1978; Mayr 1982).

[10] "*Suppose, for example, that a group of distinguished families possess potential or actual versatility to the extent of being able successfully to fill the role, either of a landed gentleman administering his estates, or of a soldier. A is the eldest son, and stays at home; his brother B goes to the wars; then so long as A has some eight children, it does not matter, genetically, if B gets killed, or dies childless, there will be nephews to fill his place*" (Fisher 1914, p. 315).

shift in focus from individuals to genes, there was also a drastic reconceptualization of the very phenomena at the heart of scientific enquiry: "*evolution is a change in the genetic composition of populations*" (Dobzhansky 1937, p. 11).

In this way, *The Modern Synthesis* is actually a rather curious text: on the one hand it was revolutionary, but on the other it was also incredibly dated. For example, Huxley explains natural selection in the form that Darwin presented in *The Origin of Species—i.e.* in the form that was not watertight and hence experienced an 'eclipse'. Huxley thought that population genetics vindicated Darwin's argument, when instead it radically transformed it. For this reason, Peter Medawar is reported to have remarked after Huxley delivered a talk: "*The trouble with Julian* [Huxley] *is that he really doesn't* understand *evolution*" (Dawkins 2013, p. 269). For this reason, the constructed history of this time by mainstream evolutionary biologists gives much more attention to a later work by George Williams (1966), called *Adaptation and Natural Selection.*[11] This book was enormously influential in firmly placing the externalist approach to adaptation from Paley, which is so prominent in Darwin's description of evolution by natural selection, in centre-stage.[12]

Williams brings a new philosophical rigour to the concept of adaptation, only licensing its use under restrictive circumstances: describing a trait as an adaptation is a specific hypothesis about what trait is being considered, what functions that trait serves and what aspect of the environment drives the trait's selection.[13] Central to this reinstatement of Darwin's question of adaptation was an abandonment the individual-centric description of evolution by natural selection on traits and a rehousing of the basic idea within a gene-centric framework.[14] This may seem a little confusing, because Williams is also looked back to for asserting individual selection over group selection, but this assertion is made because Williams is thinking about those entities as genetically-accounted.[10] This line of reasoning was taken to its logical extreme by Richard Dawkins (1976) in *The Selfish Gene*, who extolled Williams' gene-centric approach to evolution with great flare.[15] Within Dawkins' 'seductive' description,

---

[11] "*Williams'* [1966] *shift in emphasis from individuals to genes went almost unnoticed. His interpretation has not only peacefully coexisted with the synthetic theory for two decades, but has also been typically regarded as a brilliant defence of it. Williams' genic selection, however, has taken on a new-found importance. When genic selection was contrasted with selection on populations, it drew little attention, as most people mentally equated genic selection with individual selection. However, with the publication of Richard Dawkins's The Selfish Gene, genic selection was pitted against individual selection. In Dawkins's work, the significance of Williams' seemingly subtle shift in emphasis became focused and clearly associated with a fundamental shift in the language of evolution*" (Buss 1987, p. 175).

[12] "*I hope that this book will help to purge biology of what I regard as unnecessary distractions that impede the progress of evolutionary theory and the development of a disciplined science for analysing adaptation*" (Williams 1966, p. 4).

[13] "*The decision as to the purpose of a mechanism must be based on an examination of the machinery and an argument as to the appropriateness of the means to the end*" (Williams 1966, p. 12).

[14] "*The natural selection of phenotypes cannot in itself produce cumulative change, because phenotypes are extremely temporary manifestations*" (Williams 1966, p. 23).

[15] "*Was there to be any end to the gradual improvement in the techniques and artifices used by the replicators to ensure their own continuation in the world? There would be plenty of time*

there is a definite hardening of what evolutionary biology is all about. When individuals are viewed as 'throwaway survival machines', attention necessarily refocuses on what is more permanent—the genes. But, one might think, surely there are other entities that could have enough permanence to also be an important part of evolutionary change? Dawkins gives a thorough exhibition of this point to discuss why genes take centre-stage: "*What, after all, is so special about genes? The answer is that they are replicators*" (Dawkins 1976, p. 191). Thus, whilst biologists might say 'genes', it is really replicators that are at the heart of evolutionary explanations. As Dawkins described, replicators have a degree of permanence unlike any other biological entities because they have the stability, fecundity and fidelity (of replication) to survive on evolutionary timescales. Whilst Dawkins does flirt with the idea of a second replicator within human culture (coining the term 'meme'), he does not think that biological evolution is impacted by any replicator other than genes[16]—and mainstream thought still concurs with this opinion.[17] The hard-won replicator perspective on evolution is widely celebrated because it makes us think clearly about 'what' is being selected.[18]

So, what is evolutionary biology all about? In the broadest sense, it is about how living things change over the generations. But, for the most part, this line of enquiry is about the genetics of adaptation: why we see one trait rather than another, from the

---

for improvement. What weird engines of self-preservation would the millennia bring forth? Four thousand million years on, what was to be the fate of the ancient replicators? They did not die out, for they are past masters of the survival arts. But do not look for them floating loose in the sea; they gave up that cavalier freedom long ago. Now they swarm in huge colonies, safe inside gigantic lumbering robots, sealed off from the outside world, communicating with it by tortuous indirect routes, manipulating it by remote control. They are in you and in me; they created us, body and mind; and their preservation is the ultimate rationale for our existence. They have come a long way, those replicators. Now they go by the name of genes, and we are their survival machines" (Dawkins 1976, p. 19–20).

[16] "*My primary intention [by introducing memes], however, was not to make a contribution to the theory of human culture, but to downplay the gene as the only conceivable replicator that might lie at the root of a Darwinian process. I was trying to push 'Universal Darwinism' (the title of a later paper, based on my lecture to the 1982 conference commemorating Darwin's death). Nevertheless I am delighted that… others have run, so productively, with the meme ball*" (Dawkins 2013, p. 280).

[17] That cultural variants are not viewed as replicators is not universally agreed upon, but the argument on both sides has become dominated by what words imply. For example, those in favour of memes (*e.g.* Dennett 2017) cite those who are not in favour as supporting their argument (*e.g.* Richerson and Boyd 2005)—despite explicitly rejecting the term 'meme' because it implies that cultural variants are replicators. In general, I would favour the approach of Richerson and Boyd's perspective because it explicitly acknowledges that the mechanism of cultural inheritance is critical to exactly what is being preserved (and does not loosely apply the replicator concept in the absence of a clear understanding of what is replicated).

[18] "*If a genetic change that lengthens the bone also curves the eyebrow, then our adaptive explanation should recognise that; we should be interested in the genetic differences that give rise not merely to differences in toe-length but to differences in toe-length-plus-eyebrow-shape, even if eyebrow shape should turn out to be selectively neutral. This is an answer that would not have been obvious to the organism-centred view of classical Darwinism but comes readily to a theory that is gene-centred*" (Cronin 1991, p. 107).

externalist perspective of the features of the environment that lead to the selection of some genetic variants rather than others.

1.2 What Is Evolutionary Biology's Concept of Agency?

I have set evolutionary biology firmly on the philosophical foundations of British empiricism. In this light, unsurprisingly, evolutionary biology's concept of agency is very much in-keeping with this tradition[19] in viewing agency as a metaphor—not a 'fact of nature'. Consequently, there is no problem in talking about genes, individuals or groups as agents, but there is a need to be disciplined. Agency is a useful way of talking about entities that take decisions, so can in principle be applied to many biological entities. But, as Williams (1966) argues in *Adaptation and Natural Selection*, we should not confuse scenarios where a single or multiple agencies are at work because the outcome can be very different (which was the cause of Wynne-Edwards' group selection controversy).

Evolutionary biologists tend to use this concept of agency in very loose ways. For example, although I have stated that agency can be a useful way of talking about genes, individuals or groups, in each case most evolutionary biologists discuss individuals and groups as genetically-accounted (*i.e.* as collections of genes). Because natural selection acts on phenotypes and acts by changing genotypes, evolutionary explanations often focus on a genotypic change but explain it via the phenotypic consequences of competing genotypes. The blurring of this replicator-vehicle distinction is something of a bad habit, but it can make arguments much easier to follow by observing the convention that the 'individual' refers to whatever genotypes of that individual are currently relevant. I would also add, here, that there is now a thriving research tradition stemming from Leo Buss's (1987) *The Evolution of Individuality* where the coherence of biological entities as discrete individuals is understood as a derived trait along a continuum of individuation at different phenotypic (or vehicular) levels across the diversity of current lifeforms (see also Maynard Smith and Szathmáry (1995) *The Major Transitions in Evolution*).

One may wonder, how does this concept of agency gel with common-sense notions like free will? When *The Selfish Gene* was published, I think that the philosophical

---

[19] "*For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe any thing but the perception. When my perceptions are removed for any time, as by sound sleep, so long am I insensible of myself, and may truly be said not to exist. And were all my perceptions removed by death, and could I neither think, nor feel, nor see, nor love, nor hate, after the dissolution of my body, I should be entirely annihilated, nor do I conceive what is further requisite to make me a perfect nonentity. If any one, upon serious and unprejudiced reflection, thinks he has a different notion of himself, I must confess I can reason no longer with him. All I can allow him is, that he may be in the right as well as I, and that we are essentially different in this particular. He may, perhaps, perceive something simple and continued, which he calls himself; though I am certain there is no such principle in me*" (Hume 1738, p. 134).

impact of the idea that your genes may have contrary interests to your own presented the ‘self’ as something of that old Cartesian duality in suggesting a ‘ghost in the survival machine’. This is not what evolutionary biology implies. Instead, following empiricism, the general view is one of compatibilism between causal determinism and human freedom[20]: all events are seen as part of chains of cause and effect, irrespective of whether or not those causes or effects are necessarily observable, and human freedom is viewed as a subjective statement about our incomprehension of how our own causal mechanisms work (rather than inviting speculation on whether or not there are unaccounted supernatural sources of causation).

1.3 Why Does Evolutionary Biology Give a Privileged Role for DNA?

I have stated, in broad terms, that evolutionary biology is about how living things change over generations, and this means understanding why we see one trait rather than another, from the externalist perspective of features of the environment that lead to the selection of some genetic variants rather than others. So, when thinking about the claim that evolutionary biology gives a privileged role to DNA, there are two basic responses here. The first is to deny that evolutionary biology does give a privileged role to DNA in its explanations of how traits change. I think it would be possible to argue that this misunderstands the way in which genetic explanations of adaptation draw links between the environment and DNA, as a hypothesis that connects some external feature of the environment with some internal feature of individuals. The second is to accept that there is a kind of privilege at work, which is afforded to replicators. As the only widely-accepted replicator is the gene (*i.e.* DNA), mainstream evolutionary biology is principally about the genetics of adaptation. Most evolutionary biologists would launch into the first response, but what would follow would be a rather dull, long-winded case-by-case exposition of paradigmatic examples of how evolutionary biology asks a question and finds its answer in a genetic

---

[20] “*There is a doctrine about the nature and place of the mind which is prevalent among theorists, to which most philosophers, psychologists and religious teachers subscribe with minor reservations.… The official doctrine, which hails chiefly from Descartes, is something like this. With the doubtful exceptions of the mentally-incompetent and infants in arms, every human being has both a body and a mind. Some would prefer to say that every human being is both a body and a mind. The body and the mind are ordinarily harnessed together, but after the death of the body the mind may continue to exist and function. Human bodies are in space and are subject to mechanical laws which govern all other bodies in space. … But minds are not in space, nor are their operations subject to mechanical laws. … Such in outline is the official theory. I shall often speak of it, with deliberate abusiveness, as “the dogma of the Ghost in the Machine.” I hope to prove that it is entirely false, and false not in detail but in principle. It is not merely an assemblage of particular mistakes. It is one big mistake and a mistake of a special kind. It is, namely, a category mistake. … [p. 66] … In short, then, the doctrine of volitions is a causal hypothesis, adopted because it was wrongly supposed that the question, ‘What makes a bodily movement voluntary?’ was a causal question*” (Ryle 1949, p. 17 and p. 66 as marked).

difference. But I am not really sure that gives a serious treatment of the criticism, which I think is less about paradigmatic examples and more about the way in which the genetic focus of research can *by assumption* exclude alternative (and interesting) sources of explanation from the enquiry. In this way, I am going to only address this second response.

Claims of replicators that are built of other materials than DNA are controversial, but they are 'in the air' at the moment with the rise of epigenetics. These are not woolly suggestions of 'memes' or suchlike, which have dropped out of favour because it is not clear that the study of 'cultural variants' really gains very much from the analogy to genes (because the mechanism of genetic replication is nothing like how organisms learn).[15] This is actually a critical point: the 'eclipse of Darwinism' was made possible by the fact that Darwin's argument in the absence of a mechanism of inheritance was not guaranteed to be correct. For a non-DNA replicator, the mechanism of inheritance would have to be known for it to be more than an interesting speculation—and perhaps some epigenetic systems are sufficiently well-characterised to be worthy of exploration.

What would need to be demonstrated to evidence a non-DNA replicator? I don't view this as a systematic answer, but there would (at least) need to be a clear demonstration that the candidate biochemical was capable of traits that are independent from variation in the DNA-replicator. Consequently, instances where a candidate biochemical is inherited but not replicated would be insufficient. These might include, for example, regulatory biochemicals that are given by a mother to her unborn infant during pregnancy to 'prime' that individual for the environment they are about the experience. Even if these molecules were inherited across multiple generations, they would only survive on evolutionary timescales if they were being replicated. Regeneration of a regulator does not count as replication, because it is still under the control of the DNA-replicator. That is not to say that I, as an evolutionary biologist, do not find this fascinating, but I would tend to view priming as interesting from a different perspective. The question that interests me is about the selection on the genetic variant that enables priming rather than whether or not individuals have a primed phenotype. To date, and to my knowledge, there are not epigenetic molecules that are anything more than inherited-regulatory molecules that act as primers.[21]

For those interested in epigenetic replicators, I think there is one consideration that is always worth bearing in mind. Even if a non-DNA replicator were discovered, which I would keep an open mind toward: how would evolutionary biology change were a non-DNA replicator discovered? I would suggest, not very much. The vast majority of genetic explanations of adaptation would still hold, because the vast majority rest on the experimental manipulation of individuals with different genetic variants in order to confirm what feature of the environment selects a particular genetic variant (and its associated traits). Given this, genetic change must be the predominant explanation of how living things evolve. However, a non-DNA replicator would introduce a new dimension for evolutionary biology. Just as the quirks of the genetic mechanism influence how traits evolve, quirks of the new replicator's

---

[21] There is a thorough Neo-Darwinian discussion of epigenetics in Haig (2007).

mechanism would presumably do likewise. And, I would suppose, there would also be room for conflict between types of replicator. In the history of life on earth, it is generally thought that there have been other types of replicator and that the genetic code (built on DNA) has selectively out-competed other systems because it is a good medium of replication.[22] But, it is not widely held that other types of replicator beyond those built from nucleic acids are important for the 3.8 billion years of life on earth that we currently know about. And, of course, if there were convincing evidence to contrary, opinions would change.

So, why does evolutionary biology give a privileged role to DNA? Given that evolutionary biology is trying to explain why traits change over time in one way rather than another, evolutionary biology privileges DNA in its explanations because the only genes (*i.e.* DNA-replicators) can persist on the relevant evolutionary timescales.

2 Why Is Popper's 'Active Darwinism' Problematic?

I have presented an explanation of why evolutionary biology is set up the way it is, and now I want to turn to the alternative concept that was advocated by Popper and pushed for with renewed vision by Noble.

2.1 What Is Popper's Reading of Evolutionary Biology?

Popper's earliest evolutionary ideas were first expressed in *The Poverty of Historicism* (1957) and came to the fore in *The Logic of Scientific Discovery* (1959) where the growth of knowledge was described as a process of cumulative error elimination. But Popper did not see an immediate parallel with his theory of scientific progress and evolutionary biology. By the time of his *Intellectual Autobiography* (1974), Popper started to make these connections but was rather wary of 'Darwinism': "*I have come to the conclusion that Darwinism is not a testable scientific theory, but a* metaphysical research programme—*a possible framework for testable scientific theories*" (p. 134). For him, this rests of the premise that: "*Darwinism does not really* predict *the evolution of variety. It therefore cannot really* explain *it. At best, it can predict the evolution of variety under "favourable conditions". But it is hardly possible to describe in general terms what favourable conditions are—except that, in their presence, a variety of forms will emerge*" (p. 136). It is nonetheless clear that evolutionary biology is especially problematic for his understanding of scientific progress, but he firmly states: "*And yet, the theory is invaluable*" (p. 137). In this early interaction, I think we can see how Popper is seeking to isolate a specific strand

---

[22] These ideas are put forward by Cairns-Smith (1982), and enthusiastically discussed by Dawkins (1986). A more modern treatment is given by the seminal Maynard Smith and Szathmary (1995).

of evolutionary biology as 'Darwinism', as opposed to a more general evolutionary approach which he sees himself as a contributor toward.

Following on from ideas developed in the Spencer Lecture (1961) entitled *Evolution and the tree of knowledge*—which was the basis of a chapter in *Objective Knowledge* (1972), Popper controversially expresses dissatisfaction that evolutionary biology can adequately explain cumulative adaptation, following the suggestions of others that there must be 'orthogenetic trends' to funnel variation in specific directions. But at the time of completing *Objective Knowledge* (1972), unambiguously stated that the "*Neo-Darwinist theory of evolution is assumed*" (p. 242), and he went on to elucidate twelve theses on which evolutionary theory rests, which can be broadly summarised en-masse as applying his thinking of cumulative error elimination within scientific progress to nature. I would suggest that he was starting to see his ideas on the growth of knowledge within the broader context of evolutionary thought (*i.e.* seeing epistemology as an evolutionary science), alongside a long-standing unease with something in the contemporary science. Popper's insistence that there is a common mechanism at work within scientific progress in knowledge and adaptive evolution in nature[23] was in tune with the *zeitgeist*, where there was enthusiasm for 'Universal Darwinism'[24] and the broader development of 'evolutionary' subdisciplines (most notably) in economics, computer science and psychology.

Although it is not clear at exactly what stage Popper read various works of Samuel Butler, especially *Evolution: Old and New* (1879) where the basic distinction between 'active' and 'passive' Darwinism is first made,[25] Popper acknowledges a debt toward him in his *Intellectual Autobiography* (1974) whilst expressing a general disdain

---

[23] "*In my opinion, passive Darwinism turns out, when confronted by active Darwinism, to be a mistaken interpretation of the process of adaptation. Adaptation is, I suggest, essentially a trial and error learning process that extends over many generations. … [p. 121] … We should regard the whole of evolution as a huge learning process going in all sorts of directions and specialisations*" (Popper 1986, in Niemann 2014, p. 120 and p. 121).

[24] "*My general point is that there is one limiting constraint upon all speculations about life in the universe. If a life-form displays adaptive complexity, it must possess an evolution mechanism capable of generating adaptive complexity. However diverse evolutionary mechanisms may be, if there is no other generalization that can be made about life all around the Universe, I am betting it will always be recognizable as Darwinian life*" (Dawkins 1983, in Bendall 1983, p. 423).

[25] "*In like manner we say that the designer of all organisms is so incorporate with the organisms themselves—so lives, moves, and has its being in those organisms, and is so one with them—they in it, and it in them—that it is more consistent with reason and the common use of words to see the designer of each living form in the living form itself, than to look for its designer in some other place or person. Thus we have a third alternative presented to us. Mr. Charles Darwin and his followers deny design, as having any appreciable share in the formation of organism at all. Paley and the theologians insist on design, but upon a designer outside the universe and the organism. The third opinion is that suggested in the first instance, and carried out to a very high degree of development by Buffon. It was improved, and, indeed, made almost perfect by Dr. Erasmus Darwin, but too much neglected by him after he had put it forward. It was borrowed, as I think we may say with some confidence, from Dr. Darwin by Lamarck*, and was followed up by him ardently thenceforth, during the remainder of his life, though somewhat less perfectly comprehended by him than it had been by Dr. Darwin. It is that the design which has designed organisms, has resided with, and been embodied in, the organisms themselves" (Butler 1879, p. 24–33).

for other ‘evolutionary philosophers’. Many of the concepts of Popper’s Medawar Lecture (1986, published in Niemann 2014) are within Butler’s *Evolution: Old and New*, but I suspect that Popper arrived at Butler’s perspective semi-independently in marrying together a dissatisfaction with a specific strand of evolutionary thought and the comparisons with (and generalisation of) his ideas on the growth of the scientific knowledge.

Within the Medawar Lecture, Popper clearly expressed an understanding of the essential aim of what evolutionary biology was about,[26] but disagreed with much of the language in which ideas are presented. This disagreement led him to discuss natural and sexual selection as competing theories,[27] when most evolutionary biologists would the latter as a subcategory of the former. Popper viewed the role of organisms’ preferences for choosing their own niche to be broadly ignored with contemporary evolutionary theory,[28] and consequently asserted a much greater role for problem-solving (*i.e.* learning) in the general picture of how organisms are selected. Nevertheless, I think is important to remember is that Popper was contrasting two forms of Darwinism, in that he wasn’t suggesting that Darwin’s research tradition was ‘wrong’—or advocating some alternative like Lamarckism.[29] Instead, I think his aims in the Medawar Lecture were more in the vein of stating some things that appeared odd within the framework of contemporary evolutionary theory from the opinion of an outsider. And, in short, what struck him as odd was evolutionary biology’s concept of agency.

---

[26] “*My problem exists because some excellent Darwinists even believe that evolution can be fully explained by only two things: (1) The variability of the genome whose variations are obviously a matter of chance are completely independent of the organisms’ activities and preferences; and (2) The physical environment, where ‘physical’ may include, of course, the physical presence of other organisms*” (Popper 1986, in Niemann 2014, p. 119).

[27] “*Darwin, as you all know, believed in sexual selection. And he believed that sexual selection was a kind of natural selection. But this is only if we take the niche of the male, to which the female belongs, as the niche that is here important. It can be easily show, all of you can think of this when you go home, that if we take a niche that covers both male and female, then sexual selection is a refutation of natural selection. So it depends on the concept of niches whether sexual selection fits into the scheme of natural selection or refutes it. If you take the niche of the male, then the female is part of the niche and the male must please the female by such things as tail or horns, or I do not know what, which may not be very useful for natural selection. But if you take the niche for male and female together, then most of the examples of sexual selection are a worsening of adjustment, of the adaptation, to this common niche. They are an improvement of adaptation to the niche of the male and a worsening of adaptation to the common niche of male and female*” (Popper 1986, in Niemann 2014, p. 127–128).

[28] “*One of my assertions is that the preference for better niches is the main thing that leads to Darwinian evolution. The organisms are active. They search a better niche. And then this niche, this environment, ensures somehow that the better adapted organisms leave more offspring. And in this manner we get specialisation and more adaptation*” (Popper 1986, in Niemann 2014, p. 122).

[29] “*I do not defend Lamarckism as it is today called, that is to say, the inheritance of acquired properties*” (Popper 1986, in Niemann 2014, p. 127–128).

2.2 What Is Popper's Concept of Agency?

Popper defends a common-sense notion of agency, which he exclusively attributes to organisms, based on the fact that you have free will in the very literal sense that you have real choice that is not determined by any prior events (*contra* causal determinism)—but it is fair to say that his views here are quite hard to discern. Prior to the Medawar Lecture, Popper takes the view that human agency is somewhat exceptional in contrast to other animals' agency—though we share some basic features.[30] But, whilst human knowledge is primarily learnt about their world and consequently agency develop as an ability to make choices,[31] animal knowledge is primarily genetic—having been acquired through natural selection.[32] Popper clearly had an uneasy relationship with what he referred to as either 'materialism'[33] or 'determinism'[34]—but what I will refer to as (British) empiricism. My reading is that Popper struggled to work out what kind of a claim he was wanting to make: was free will a claim about our imperfect understanding of human behaviour or a claim about how humans are? Here, Popper sided with rationalism rather than empiricism, to assert that consciousness (and hence free will) is an undeniable objective fact.[35] To a rationalist, consciousness is the first and most undeniably true fact of existence because it does not rest on anything other than introspection. But, following

---

[30] "*I assert that every animal is born with expectations or anticipations, which could be framed as hypotheses; a kind of hypothetical knowledge. And I assert that we have, in this sense, some degree of inborn knowledge from which we may begin, even though it may be quite unreliable. This inborn knowledge, these inborn expectations, will, if disappointed create our first problems; and the ensuing growth of our knowledge may therefore be described as consisting throughout of corrections and modifications of previous knowledge*" (Popper 1972, p. 258–259).

[31] "*It seems to me of considerable importance that we are not born as selves, but that we have to learn that we are selves; in fact we have to learn to be selves. … [by] developing theories about ourselves*" (Popper and Eccles 1977, p. 109).

[32] "The believer—whether animal or man—perishes with his false beliefs." (Popper 1972, p. 122)

[33] "*I do not claim that I have refuted materialism. But I think that I have shown that materialism has no right to claim that it can be supported by rational argument—argument that is rational by logical principles. Materialism may be true, but it is incompatible with rationalism, with the acceptance of the standards of critical arguments; for these standards appear from the materialist point of view as an illusion, or at least as an ideology*" (Popper and Eccles 1977, p. 81).

[34] "Indeterminism—*or more precisely, physical indeterminism—is merely the doctrine that not all events in the physical world are predetermined with absolute precision, in all their infinitesimal details*" (Popper 1972, p. 220).

[35] "*We have to assume, difficult as this may be, that it [consciousness] is a product of evolution, of natural selection. Although this might constitute a programme for a reduction, it is not itself a reduction, and the situation for the reductionist looks somewhat desperate; which explains why reductionists have either adopted the hypothesis of panpsychism or why, more recently, they have denied the existence of consciousness (the consciousness say, of toothache) altogether. Though this behaviourist philosophy is quite fashionable at present, a theory of the nonexistence of consciousness cannot be taken any more seriously, I suggest, than a theory of the nonexistence of matter. Both theories 'solve' the problem of the relationship between body and mind. The solution is in both cases a radical simplification: it is the denial either of body or of mind*" (Popper 1974, p. 272–273).

Searle's (1999) terminology, Popper seems to conflate statements that are epistemically objective (*i.e.* claims about what is from my perspective) with statements that are ontologically objective (*i.e.* claims about what is). The former are dependent on current evidence, but the latter are not. In contrast with the empiricist tradition, free will is an epistemically objective 'illusion'[19]—but that is not to say that behavioural science would ever have enough knowledge to predict human behaviour with any reasonable accuracy. I have often wondered whether Popper's view was influenced by his point in history, where he had seen the damage that could be done by entertaining a nihilistic view of the objective world.[36] I might also add that I have always been baffled why indeterminism might somehow make room for free will in the objective world (what Popper called World 1), when its behaviours remain statistically definite.

2.3 Why Is Popper's 'Active Darwinism' Problematic?

I do not think there is any other way to construe my reading of Popper's division of 'active' and 'passive' Darwinism: it is a false dichotomy. The thing that really makes me firm about this conclusion is Popper's treatment of sexual selection, which he argues contradicts natural selection.[23] I understand what he means, namely that what makes an individual adapted to survival can differ from what makes an individual adapted for reproduction—which is not a controversial statement. But natural selection is generally used as the overarching idea of any type of selection, of which sexual selection, kin selection, fecundity selection, mortality selection etc. are subtypes. More to the point, I see no deficit in the current research paradigm stemming from Darwin, who gave us both the concepts of natural and sexual selection. I reject Popper's assessment of theory only treating the male's choice (or niche), which probably stems from Popper being unaware of traits that are associated with female choice—but nevertheless if he was aware of these cases then he glossed over them as he was running out of time at the end of the lecture. In this way, much of this disagreement about sexual and natural selection must surely reflect a problem of language, as Popper was clearly using terms in a different way than evolutionary biologists' do. I wonder how much of the general idea of 'active Darwinism' is of the same flavour, but I am not going to focus on this exposition because it seems rather dull.

At a deeper level, the problem with 'active Darwinism'—in as far as there is one that extends beyond a rephrasing of the ideas in 'passive Darwinism'—relates to Popper's discussion of how organism's choices impact how they evolve. I do not

---

[36] "*Compton describes here [in a preceding quote] what I shall call 'the nightmare of the physical determinist'. A deterministic physical clockwork mechanism is, above all, completely self-contained: in the perfect deterministic physical world there is simply no room for any outside intervention. Everything that happens in such a world is physically predetermined, including all our movements and therefore all our actions. Thus all our thoughts, feelings, and efforts can have no practical influence upon what happens in the physical world: they are, if not mere illusions, at best superfluous byproducts ('epiphenomena') of physical events*" (Popper 1972, p. 217).

think that there is any disagreement that, say, the sexual preferences of organisms can be important in determining how evolution proceeds. I think there is room to doubt two things: first is the generality with which this applies, and second is the essentiality to a general explanation of natural selection.

Popper is right to assert that many organisms have preferences that change how they interact with the environment, and consequently how selection acts on them; but when Popper encounters this, he asks "How do those preferences impact evolution?" when an evolutionary biologist would ask "Why are those preferences adaptive?". If those preferences were arrived at randomly, they would be of little interest to me as an evolutionary biologist because they would not be adaptive. Preferences are only going to be adaptations if they have the ability to be passed on in the longer-term (*i.e.* over many generations), which would need them to be produced by replicators. So, the fact that preferences change evolution is point of agreement, but Popper has inverted the causality to suggest the preference evolves before the gene that enables an individual to express that preference. Popper gives little indication about where this preference might come from beyond 'active problem-solving', which both Popper and his commentators have likened to a Baldwin effect, where learnt preferences (*i.e.* non-genetic adaptations) impact genetic evolution. As Popper seems to be aware, there is nothing incompatible between the Baldwin effect and what he calls 'passive Darwinism'—and the Baldwin effect is even incorporated into Huxley's *The Modern Synthesis*. The difference is more that Popper assumes that the Baldwin effect is the 'general case' and other cases are the exception (hence favouring the phrasing of his active Darwinism), whilst evolutionary biologists assume the opposite. In defence of the position of most evolutionary biologists, I could now launch into a set of evidence that not many organisms (if not only one) are capable of generalised learning in such a way that their learnt preferences meaningfully impact their genetic evolution in order to show that most lifeforms evolve in a much more 'passive' way that Popper supposed.[37] But, for me, the crux is really settled by my second point.

---

[37] To push this point further, in the Medawar Lecture, I think it is revealing that Popper spends his term advocating 'active' Darwinism in the discussion of animals *only*—and I would read him more specifically as talking about vertebrates because only they have a kind of generalised learning because of their centralised nervous system (in a way that makes individual capable of expressing its own unique personality). Further, Popper gives little consideration of the animals, plants and micro-organisms that form sedentary (or sessile) individuals that do not have much control over their environment. I do not mean to imply that they cannot engage in 'niche construction', but sedentary species clearly must have a restricted ability to do so in comparison to motile species. The fact that there are degrees to which Popper's 'active Darwinism' might be better for understanding some species over others is very different from disputing the 'general case', as Popper does. Additionally, Popper gives no discussion of selfish genetic elements, intragenomic conflict and horizontal gene transfer and other phenomena of living things that undermine the importance of individuals as coherent/unified learning agents. From an empiricist perspective, I think that Popper falls into a rationalist trap, which was eloquently stated by Hume: "*What peculiar privilege has this little agitation of the brain which we call thought, that we must thus make it the model of the whole universe?*" (Hume 1779, p. 134). In this reading, Popper applies the structure of his own way of thinking to the objective world as if the objective world had the same structure; hence why Popper thought it was legitimate to draw parallels between the growth of scientific knowledge and evolution

Organisms do not need to be active problem-solvers for them to evolve by natural selection, and so problem-solving does not really explain adaptation in general terms. It may well be true that generalised learning has a much greater role in evolution than most evolutionary biologists tend to think, but natural selection would work on entities that are incapable of learning. Indeed, evolutionary biologists tend to mostly work with genes, which may react to different environments in different ways (which we can understand as a probabilistic 'reaction norm') but are fundamentally inert chemicals that do not change their own base composition. Instead, environmental factors may cause them to mutate as they are passed down the generations, and therefore any adaptations that they contribute toward are only the result of natural selection. On this point, it can be useful to follow evolutionary epistemology's portioning of an adaptation into components of instruction and selection (Plotkin 1994). The basic idea here is that adaptation can come about through two basic sources: instruction refers to following some 'rules', and selection refers to environmental feedback on blindly-generated variation.[38] Classically, these two sources of adaptation can be thought of as extremes on a continuum between rules uniquely specifying a single adaptive variant and the generation of multiple variants that are then whittled down to a single adaptive variant. A preference is an instruction, but the question is how any adaptive properties were arrived at. If the preference is innate, then it was arrived at through selection on genotypic variants. If the preference is learnt, then it was arrived at through instruction by some phenotypic set of rules. The continuum perspective masks that those phenotypic rules are only going to successfully lead to adaptation if they are the result of a selective process (*i.e.* by selection on genotypic variants that specify learning rules). Therefore no matter how you look at it, in the ultimate sense, adaptation is arrived at through natural selection somewhere in the system, but it doesn't necessarily have to be natural selection in a straight-forward manner. I would also say, tying this back to biology, adaptation is arrived at because of the natural selection of genotypic variants that underlie behaviour, or the learning rules that govern behaviour. In this way, it is natural selection not problem-solving (*i.e.* learning) that is essential to adaptation.[39]

---

by natural selection. In other words, I think that Popper over-states the degree to which organisms choose their environment because his philosophy lends him toward being anthropocentric.

[38] The critical feature of selection is sometimes misconstrued (as Noble does so), and so I will clarify. Although the paradigmatic selection process would be random variation, a bias in the process of mutation does not matter in as far as the bias does not influence the outcome of the selective process—which is to do with a feature of the environment that does the selecting. This is why the word 'blind' is often preferred to 'random'.

[39] As an aside, I think it would be possible to make the reverse argument that instruction underlies selection, but to do so would require the physical laws of the universe to be construed as 'instructions'. In this implicitly causally deterministic framework, natural selection would proceed from the physical laws of the universe because the physical laws permit selection to operate. This argument is not totally vacuous, hence why natural selection can be simulated in a computer that operates by a series of instructions. However, I would argue that this argument alienates an important aspect of natural selection, which is the medium of the replicator. The mechanics of genome replication has a huge impact on the direction of the evolutionary change resulting from selection—and the degree to which different kinds of traits can be more or less adaptive, so I am not sure how useful

Perhaps one could argue that this restricts the scope of evolutionary biology's explanation of adaptation, but I think it clarifies something very important. If there is any adaptation as a result of learning, that adaptation is either the result of the natural selection of genes governing how behaviour changes in response to some feature of the environment or the result of a secondary phenotypic process of selection that is enabled (but not directed) by genes. The second case, we might consider as 'open-ended' or 'generalised' learning, though of course how open-ended it is depends on the system's constraints (just as modern genes are constrained by protein biochemistry). To explain a trait as an adaptation would require an intimate knowledge of the way selection works in that system (just as natural selection only made sense given genetics). In the context of genes, this is often described as suggesting that genetic 'constraints' are an important part of adaptive explanation because of their creative role in how selection works.[40] What little we do know about learning systems is that they vary across the diversity of life, and so the constraints in these systems are never going to be universally shared (unlike for the genetic code, which is pretty much universal). The point I want to make here is this: even if evolutionary biologists were interested in learnt adaptations, we would explain them with the same externalist mindset as we apply to genetic adaptations. In this way, we would still direct focus toward how features of the external environment cause features of the internal structure that we see, rather than really treating individuals as active problem-solvers.

3 What Was Popper's Criticism Really About?

This brings me on to my final set of thoughts. I find it hard to believe that Popper cannot have considered most of the arguments that I have just raised both for evolutionary biology's framework and against his suggested alternative. The question is, why did he continued anyway? I don't think that he was really trying to revolutionise evolutionary biology by unveiling some fatal flaw in contemporary research. Instead, I think he was trying to push the analogy of evolutionary change as a learning process in order to expose something odd about the way evolutionary biology conceptualises agency. I think Popper rightly sensed the externalist tendency of evolutionary biology

---

this perspective is. Further, given its causal determinacy, I am not sure how much is reinforces the active problem-solving perspective of organisms.

[40] "*It is common to think of constraints in a negative fashion – as preventing things from happening, and thereby reducing the variety found in nature. But if the process of producing variation is open-ended, the introduction of constraints can channel the variation, and by directing it, produce much further or deeper exploration in a given direction than would otherwise be possible. Constraints can thus play a creative and, in one sense, ultimately progressive role. This is a deep truth, not only about evolution, but about problem-solving and exploration in general. It is why Darwin was right in 1859 when he saw natural selection as a creative force, and why his critics who saw selection only as playing a negative role by eliminating variety were wrong*" (Wimsatt and Schank 1988, in Nitecki 1988, p. 235).

to explain adaptations with appeal to features of the environment, though I do not think that he recognised this externalism explicitly. Given that Popper is the champion of common-sense, I think Popper's discomfort with evolutionary biology came from his unease with the way that it flaunts agency as a metaphor—which stands very much in contradiction of his rationalist account of science as yielding objective knowledge about reality. I think that many people would instinctively sympathise with Popper's position, especially for those in contemporary thought that look to evolutionary biology as orchestrating a modern (and atheistic) Creation Myth, whereupon I think it is natural for many people to feel like there should be some greater prominence of the individual within this scientific epic. The way in which evolutionary biology asserts the insignificance of agency is omnipresent in the way in which, even when organisms are discussed, organisms tend to be talked of in terms of their genetic accounting only.

Prior to the Medawar Lecture, I have suggested that there was a tension in Popper's thoughts on evolutionary biology—hence why he both lavished it with praise and yet gave it special treatment as an inconvenient anomaly. By the time of the Medawar Lecture, I think that Popper had resolved some of this tension by asserting that agency objectively exists and is an important part of the causal structure of the objective world, rather than asserting that agency has a subjective existence as a metaphorical way of thinking about the objective world. However, I do not think that Popper critically assessed why this disagreement about agency came about. I have characterised this as Popper's favouring of rationalism over empiricism in asserting the existence of agency prior to any evidence which may suggest an alternative conclusion. Within rationalism, agency is the bedrock of all human understanding which is built from ontologically objective knowledge; within empiricism, agency is more often used as a metaphor (or 'thinking tool'), and human understanding is built from epistemically objective knowledge (which may or may not turn out to ontologically objective). In this regard, Popper's own philosophy of 'critical rationalism' is—to some extent—bridging the divide, but in other ways it is also a bridge built from one side on a rationalist foundation. Along with other scientists, evolutionary biologists tend to admire Popper as 'their' philosopher of science, in defending a common-sense world-view held by most scientists. But, in the details, I think that many scientists would defend the same world-view but from an empiricist foundation (perhaps, 'critical empiricism'?).

In this way, I think it is inaccurate for Noble and others to use Popper as someone who was 'on their side' against the views expressed in *The Modern Synthesis* because I think that Popper's complaint with evolutionary biology was a philosophical one relating to agency. Popper thought very favourably of Medawar's critique of Teilhard de Chardin's evolutionary theology, wherein scientific research was described as the 'art of the soluble'.[41] Perhaps influenced by the empiricist foundations of

---

[41] "*No scientist is admired for failing in the attempt to solve problems that lie beyond his competence. The most he can hope for is the kindly contempt earned by the Utopian politician. If politics is the art of the possible, research is surely the art of the soluble. Both are immensely practically-minded affairs*" (Medawar 1967, p. 97).

evolutionary biology, Medawar was expressing the fact that a good research scientist spends their time solving problems rather than building syntheses. Particularly in biology, a synthesis is always going to be constructed as a teaching aid for general intuition rather than as a rigorous statement of universal truth because there will always be an exception. To my mind, focusing criticism on *The Modern Synthesis* as a seminal work is a fascinating construction of the history of evolutionary theory because, as I and others have argued, it actually had very little impact compared to other contemporary works. Further, the word 'synthesis' makes it a wonderful straw-man; ecology does not have a 'modern synthesis' equivalent but is instead a looser collection of canonical concepts and so it is much harder to decry its failures in this way.

4 Conclusion

So, where does this leave us? Popper's foray into evolutionary biology is fascinating because it represents a collision of world-views. I am not sure that Popper gets everything right, and I am not sure that Noble is correct that we need to rehabilitate Popper's views of evolutionary biology—nor do I think we will ever agree on that one. But I respect that there is something non-trivial about these disagreements, which deserves further discussion. I think much of the disagreement comes from the competing treatments of agency within rationalist and empiricist traditions, and so I think that the non-trivial elements of the disagreement are philosophical in nature, rather than relating to anything that could be changed on the practical side of the established facts that either tradition could use to support their position. Perhaps by simply recognising the nature of this disagreement, a lot of misrepresented 'hot air' can be avoided.

References

Alexander, R.D.: Darwinism and human affairs. University of Washington Press (1979)

Bowler, P.J.: The eclipse of Darwinism: anti-Darwinian evolution theories in the decades around 1900. The Johns Hopkins University Press (1983)

Bowler, P.J.: Evolution: the history of an idea. University of California Press (1984)

Buss, L.W.: The evolution of individuality. Princeton University Press (1987)

Butler, S.: Evolution: old and new. Hardwicke and Bogue (1879)

Cairns-Smith, A.G.: Genetic takeover: and the mineral origins of life. Cambridge University Press (1982)

Cronin, H.: The ant and the peacock: altruism and sexual selection from Darwin to today. Cambridge University Press (1991)

Darwin, C.: On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. Murray, London (1859)

Dawkins, R.: The selfish gene. Oxford University Press (1976)

Dawkins, R.: Universal Darwinism. In: Evolution from Molecules to Men, 403–425. Cambridge University Press (1983)
Dawkins, R.: The blind watchmaker: why the evidence of evolution reveals a universe without design. Norton & Company, Inc (1986)
Dawkins, R.: An appetite for wonder: the making of a scientist. Black Swan (2013)
Dennett, D.C.: From bacteria to bach: the evolution of minds. Basic Books (2017)
Desmond, A., Moore, J.: Darwin. Michael Joseph Ltd (1991)
Dobzhansky, T.: Genetics and the origin of species. Columbia University Press (1937)
Fisher, R.A.: Some hopes of a eugenist. Eugen. Rev. **5**, 309–315 (1914)
Godfrey-Smith, P.: Complexity and the function of mind in nature. Cambridge University Press (1996)
Gould, S.J.: Ever since Darwin. Penguin (1978)
Gould, S.J.: The structure of evolutionary theory. Harvard University Press (2002)
Haig, D.: Weismann rules! OK? epigenetics and the lamarckian temptation. Biol. Philos. **22**, 415–428 (2007). https://doi.org/10.1007/s10539-006-9033-y
Hume, D.: A treatise of human nature. John Noon (1738)
Hume, D.: Dialogues concerning natural religion (1779)
Huxley, J.: Evolution: the modern synthesis. George Allen & Unwin, London (1942)
Kuhn, T.S.: The structure of scientific revolutions. University of Chicago Press (1962)
Maynard Smith, J., Szathmáry, E.: The major transitions in evolution. Bell & Bain Ltd (1995)
Mayr, E.: The growth of biological thought: diversity, evolution, and inheritance. Harvard University Press (1982)
Medawar, P.: The art of the soluble. London Methuen (1967)
Noble, D.: The music of life: biology beyond the genome. Oxford University Press (2006)
Noble, D.: Dance to the tune of life: biological relativity. Cambridge University Press (2016)
Paley, W.: Natural theology. Wilks & Taylor (1802)
Plotkin, H.: Darwin machines and the nature of knowledge. Penguin (1994)
Popper, K.: Intellectual autobiography. In: Schlipp, P.A. (ed.) The Philosophy of Karl Popper, vol. 1, 3–181. The Library of Living Philosophers Inc. (1974)
Popper, K.R.: Objective knowledge: an evolutionary approach. Oxford University Press (1972)
Popper, K.R., Eccles, J.C.: The Self and Its Brain. Springer (1977)
Popper, K.R.: Four texts by Karl Popper. In: Niemann, H.-J (ed.) Karl Popper and the Two New Secrets of Life, 115–138. Mohr Siebeck (2014)
Richerson, P.J., Boyd, R.: Not by genes alone. University of Chicago Press (2005)
Riskin, J.: The restless clock: a history of the centuries-long argument over what makes living things tick. University of Chicago Press (2016)
Ryle, G.: The concept of mind. Hutchinson & Co Ltd (1949)
Searle, J.R.: Mind, language and society. Basic Books (1999)
Simpson, G.G.: The biological nature of man. Science **152**(80), 472–478 (1966). https://doi.org/10.2307/2010269
Stott, R.: Darwin's ghosts: in search of the first evolutionists. Bloomsbury Publishing (2012)
Williams, G.C.: Adaptation and natural selection: a critique of some evolutionary thought. Princeton University Press (1966)
Wimsatt, W.C., Schank, J.C.: Two constraints on the evolution of complex adaptions and the means of their avoidance. In: Nitecki, M.H. (ed.) Evolutionary Progress, 231–275. University of Chicago Press (1988)

Philip Madgwick is a postdoctoral researcher in evolutionary biology, who contributed to this volume in the final year of his PhD as a member of the Milner Centre for Evolution at the University of Bath, UK. His PhD contributed toward the evolutionary understanding of conditional cooperation, using game theory to predict patterns of social behaviour in social amoeba as well as making contributions toward methods in game theory, the greenbeard concept and predicting the mechanisms of meiotic drivers. Philip currently works for Syngenta using population genetics to inform resistance management to combat malaria, but maintains a broad interest and actively publishes in the theory and philosophy of evolutionary biology. E-mail: philipgmadgwick@gmail.com

Popper, Darwin, and Biology

Hans-Joachim Niemann

1 Karl Popper as a Philosopher of Biology

Karl Popper has made a name for himself as a philosopher of science (*Logik der Forschung* 1934; *The Logic of Scientific Discovery*, 1959) and a philosopher of social sciences (*The Open Society*, 1945). Less well known is his third major work, a bio-philosophical one, which includes (*i*) a new interpretation of Darwin's evolution, (*ii*) ideas on the origin and the very core of life, (*iii*) evolutionary epistemology, (*iv*) the biological body-mind problem, and (*v*) what he called 'World 3', the world of language, theories and problems,[1] which he interprets biologically as the exosomatic tool of humans. In this article I discuss the first three items. (*iv*) and (*v*) are dealt with elsewhere (Niemann 2012a, b, 2018a).

2 Popper's New Interpretation of Darwinism

2.1 *Popper's Biological Starting Points*

Karl Popper had been interested in evolutionary biology since his earliest youth (Bartley 1987, 18), but it was only in his late 40s that his interest in biology took

---

[1] The Body-Mind Problem and World 3 in Popper (1969) and Popper (1977b), commented in Niemann (2012a). 'Exosomatic tools' are evolutionarily developed, out-of-body tools such as the webs of spiders, the nests of birds or the written world of man.

This article is a revised and extended translation of Niemann (2018b).

H.-J. Niemann (✉)
Reuthstrasse 14, Poxdorf 91099, Germany
e-mail: hjn@h-j-n.de

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_13

on scientific traits when he sought for solutions to the old body-mind problem, i.e. the question of how feelings and thoughts could have physical effects. He discussed such problems with the physicist Erwin Schrödinger, Nobel Prize winner and author of the well-known book *What is life?* (1944), that already included a molecular theory of genes (Popper 1974, chap. 30). Another cooperation began in 1951 with his friend, the neuroscientist and later Nobel Prize winner John Eccles. In 1977, they jointly published a new biological solution of the body-mind problem in their book *The Self and Its Brain* (see ref. Popper 1977b). With another friend, the biologist and Nobel Prize winner Peter Medawar, Popper discussed his reinterpretation of Darwin's theory of evolution. In the ninth and tenth decade of his life, his preferred discussion partner was the chemist and evolutionary experimentalist Günter Wächtershäuser who developed a meanwhile well-known theory about the origin of life from inorganic matter.[2] Popper's interest in the biological body-mind problem lasted a lifetime. He made his last suggestion when he was 92, only some weeks before his death (Popper et al. 2010), and his later so called 'two-step model' was a tentative solution of the body-mind problem—estimated, for example, by brain researcher Benjamin Libet (Niemann 2012a).

In 1961 Popper gave his first Herbert Spencer lecture on evolutionary biology titled 'Evolution and the Tree of Knowledge'. It was a significant improvement on Darwin's theory (Popper 1961), although it did not deny Darwin's explanation of evolution through variation and selection. It provided a theory of 'active Darwinism' in which the curiosity and preferences of living beings play the leading role and are the real creative elements of all evolution. According to Popper, Darwin's natural selection has only the function of correcting the mistakes of organisms, while the organisms themselves are always 'In Search of a Better World', as one of his book titles aptly describes (Popper 1992b). Only half a century after Popper's 1961 lecture, his many contributions to a reinterpretation of Darwinism are seriously discussed by evolutionary biologists (Noble 2013, 2014a, b).

Unfortunately, at first it was Popper's *criticism of Darwin* that caused quite a stir because he had characterized Darwin's theory as an example of an unscientific theory.[3] Due to his well-known *Logic of Scientific Discovery* the empirical refutability of a theory is indispensable if the theory is to be part of science (Popper 1934/1968, chap. 10). Some formulations of Darwinism do not pass this test because they cannot be falsified, for example Herbert Spencer's explanation of the origin of species with the help of the theory of 'survival of the fittest' (Spencer 1864–1867). Since the most survivable mostly survive, this theory is hardly empirically refutable, and as Popper's wisdom has it: irrefutability of theories is not a virtue, but a serious deficiency.

Nevertheless, even as an untestable theory, Popper regarded Darwinism as a fruitful 'metaphysical research programme' that one day may well attain the status of

---

[2] In his unpublished lecture, Popper (1989b) presented not his own theory, but that of Günter Wächtershäuser to a surprised audience.

[3] Popper (1972/1979), chap. 2, sect. 16; chap. 6, sect. 18. In Popper (1957, chap. 27). Darwin's evolutionary theory served as an example for failed attempts to establish laws of history.

a scientific theory (Popper 1974, chap. 37). Since 1965 at the latest, Popper considered many theses of Darwinism to be testable, such as genetic variability and natural selection: "I have changed my view on the testability and logical status of the theory of natural selection".[4]

Popper was not Darwin's opponent, but one of his admirers. He fully agreed with pre-Darwin's *theory of evolution* as well, according to which all life has developed in many millions of years from simple beginnings, and he stressed that he did so since the age of 12 (Popper 1986/2014, 118). Later he accepted Darwinism as a science, because the phrase 'survival of the fittest' can easily be avoided by a better formulation: "Organisms that are better adapted than others have a better chance of leaving offspring" (Popper 1986/2014, 121). Popper also accepted the Neo-Darwinian theories of heredity, genetic variation, and population genetics.

2.2 All Living Beings Control Their Own Evolution

'Darwinism', 'Neo-Darwinism' or the 'theory of natural selection' are the starting points for Popper's improvements. He concentrates on the two then indisputable factors that were allegedly solely responsible for evolution:

1. Variation of DNA due to random processes.
2. Natural selection.

As most biologists, Popper used the term 'natural selection' just as a vivid and useful metaphor and did not think of selection as it is done by animal breeders. He assumed that 'natural selection' is used in the sense of 'selective influence of the physical and organic environment'. In his view, expressions like Spencer's 'survival of the fittest' (Spencer 1864–1867) or Darwin's 'struggle for life'[5] are ideological expressions and not scientific ones.

Until then and even into our days, for many biologists this 'passive Darwinism', as Popper called it, was without any alternatives,—if we refrain from creationism as an alternative,[6] because it attaches no importance to scientific verifiability. Popper proposes a new, testable alternative, which he calls *active Darwinism.*

He considers the two factors (1) and (2) as misleading and untrue, because it is not true that they are the *only* causal factors of evolution. Later, Popper is to show that what is creative are neither mutations triggered by blind chance nor natural selection eliminating unsuccessful organisms. Rather, the creativity of evolutionary processes stems from the activity of living organisms.

---

[4] Popper (1977a), 144. Popper revoked his rejection of Darwinism as a science already in his (1972/1979, chap. 6) at the beginning of sect. 18, p. 241: "I blush when I have to make this confession". However, he maintained his accusation of tautology described in the previous paragraph.

[5] The 'struggle for life' is not an invention of the Darwinists or social Darwinists, but appeared, of course, already in the title of Darwin (1859).

[6] 'Creationism' is an early theory of descent based on the biblical story of creation.

It is the individuals with their peculiarities, preferences and activities that play an important role in evolution. Before Popper, others have recognized this, for example James Baldwin, after whom the 'Baldwin effect' is named (Baldwin 1896): The olm evaded its predators by retreating into dark caves. Since the genetic material differs slightly from individual to individual, natural selection preferred the ones who could perceive the environment with other sensory organs slightly better than with their eyes. Thus, the sensory properties of olms shifted from eyes to other organs. The eyes became increasingly blind.

This Baldwin effect was known for a long time, but in Darwin's theory it played only a small supporting role. Popper deserves to be known as the one who has (1) shown that the peculiarities, preferences and activities of individual organisms are not side effects but *the most important factors* of evolution. He was the first to recognize that (2) the *direction* of evolution largely depends on the activity of the organisms and that (3) the *creativity* of evolution is exclusively based on the activity of the individuals.

This theory, which Popper has advocated since his Herbert Spencer lecture (Popper 1961), is only recently gaining increasing importance among biologists of the 'Third Way of Evolution' and 'extended Darwinism'.[7] It contains many valuable arguments and explanations e.g. his 'spearhead theory' and his 'genetic dualism' which I cannot go into detail here, but for which I recommend the original literature (Popper 1961) and comments (Niemann 2013a).

2.3 Darwinism Tries to Explain Life Away

Among the Darwinists of Popper's time, but also among most of today's biologists, there is a tendency to attribute the creative power of evolution to accidental mutations and natural selection. This "Passive Darwinism is", as Popper explains, "the victim of certain pessimistic ideologies" (Popper 1986/2014, 120); above all, it falls victim to the ideology of determinism. One of Darwin's most important interpreters, the biologist Thomas Henry Huxley, was a determinist, and Popper was concerned that this ideology was still so widespread and virulent, that the distinction between active and passive Darwinism could only be understood once determinism would finally be buried.

Darwinists, Popper lamented,—and I think he meant Neo-Darwinists, as so often when he spoke of 'Darwin' in a very general way—are trying to explain life away (Popper 1970). They want to get rid of everything spiritual, such as wishes and activities of the organisms and their being oriented towards goals. They try to explain evolution only with the help of the mechanism of variation and selection. They want

---

[7] Niemann (2014a) and about 50 authors such as James Shapiro, Denis Noble, Eva Jablonka, Marion Lamb, who have contributed to the website www.thethirdwayofevolution.com for the research programme 'Neo-Darwinism needs rethinking and alternative thinking'. This is Popper's very concern since 1961. On the list are also Barbara McClintock and Conrad Waddington who had already contributed experimental preparatory work in the 1940s.

everything organic to be reduced to inorganic: Biology to biochemistry; biochemistry to chemistry; and chemistry to physics. These uncritically adopted remnants of the ideologies of materialism and determinism are what prevent them from accepting or even naming things that go beyond physics and chemistry.

2.4 Natural Selection Is Not Creative

Those who are sticking with these lines of thinking, almost deliberately ignore a whole world of extremely interesting relations, although it is not so difficult to see where the creativity of evolution comes from. I will mention here two of Popper's arguments.

A thought experiment recently discovered in the Popper archive in Klagenfurt shows that it is not possible to explain the creativity of evolution by natural selection (Popper in Niemann 2014a, appendix C). Popper constructs a world without natural selection where all species ever created still live today. Natural selection cannot have produced this diversity because it does not occur in this thought experiment. Thus, the creativity of evolution and what Darwin called the 'wonders of nature' cannot come from natural selection, nor from the struggle for survival, nor from the survival of the fittest.

Twenty years after Popper's death, experiments of the evolutionary biologist Andreas Wagner support Popper's thoughts: "Natural selection can *preserve* innovations, but it cannot create them" (Wagner 2014, Prologue).

Popper presented a second argument in his 1986 *Medawar Lecture* (Popper 1986/2014, 125). Life in its beginnings had to adapt to its environment. In some regions, life thrives easily, in others rather poorly. By the method of trial and error, the individual organisms can find better living conditions or ways of life better adapted to their environments, but it is only thanks to *creative activity* that they find unusual niches where life is easier and better than elsewhere. Thus, gentians populate high mountain areas, mice live under the ground, frogfish populate the depths of the oceans. Many organisms imaginatively change their environment or create one of their own: beavers dam rivers, birds build nests, spiders weave nets.

Those who make mistakes in their attempts to find new worlds or adapt to them are less likely to have descendants or will die. This is called 'natural selection', and before Popper, it was believed to be the creative drive in evolution that created all the species. Even Karl Popper believed this until 1961 (Popper 1961), but there is nothing creative in this process of eliminating errors through extinction. The creative part in applying the method of trial and error does not lie in error elimination, but in the attempt to try out new habitats or new ways of life. And thus, we owe the creativity of evolution almost exclusively to the fact that all living beings are problem solvers who are constantly looking for a better environment.

2.5 Creative Niche-Search Instead of Deadly Competition

The final result is that natural selection is not as important as passive Darwinism maintains. Escaping deadly errors and sheer survival are not the core of evolutionary learning. The much more important thing in the process of learning is trying out new behaviours and searching for a better world.

How the results of this search are inherited—in a Lamarckian, Darwinian, epigenetic or any other way—is irrelevant for answering the questions about creativity and the direction of evolution. Creativity does not come from blind variation and natural selection, but primarily from the activity of the organisms themselves. Equipped with curiosity, desires and preferences, organisms are constantly trying to find more pleasant environments and more suitable ways of life. Far too long, fighting and destroying competitors have been overemphasized as the biological principle of life. Popper (1986/2014, 122) claims

> that the preference for better niches is the main thing that leads to Darwinian evolution. The organisms are active. They search a better niche. And then this niche, this environment, ensures somehow that the better adapted organisms leave more offspring.

Popper speaks of the ‘niche’ in a very general sense, which also includes ways of life which allow one species to live unrivalled in the same habitat as other species: Cows and butterflies live in the same meadow but do not disturb each other. This idea of biological ‘niches’ was later generalized to ‘possibility spaces’ (Popper 1990; Niemann 2015, 2017).

The search for niches leads to specialisation. Once a new niche has been found by trial and error—as said before: by creative trials—, the task of improving the adaptation to this niche is again a matter of trial and error. So, adaptation is a twofold learning process: (1) finding a suitable niche and (2) learning to exploit its new possibilities, which means that the organism must acquire a lot of knowledge about this niche. Adaptation is not only somehow connected with knowledge, but as Popper (1986/2014, 123) explicitly points out, the two are closely connected: “I more or less equate adaptation and knowledge”.

3 Adaptation Is Active Knowledge Acquisition

3.1 Not Life, Its Adaptation Is Extremely Unlikely

Since the very beginning, all life must have been connected with the acquisition of knowledge.[8] As Popper demonstrated in another thought experiment, first life must have had big problems with adaptation, because adaptation means learning a great deal about the environment (Popper 1986/2014, 122).

---

[8] Popper (1986/2014), 123, passim; Popper (1990), no. 12–17, 36–38; (1986/2015); (1961), sect. 1.

In this thought experiment, Popper assumes that some laboratory may have succeeded in creating artificial life. Something chemical is in the test tube, it can be fed, it excretes what remains after digestion, it multiplies by division, and it never stops growing. The researchers agree that this is really about life. But something important is missing that characterizes all natural life: It is in no way adapted to its environment—which is, of course, the test tube. It is adapted to the researchers who supply it: It must be fed and protected from the outside, and its metabolic wastes must be disposed of. And as soon as things get tight in the test tube, the researchers have to make sure that the 'offspring' continue to grow up in other test tubes and are looked after there. To resemble real life, a biological machinery would have to be built into this artificial life that 'knows' everything the researchers know about the preservation of life. Among many other things life would also have to know how to find a suitable environment and how to extract from it the substances to feed the synthetic life. Life that is to survive on its own must have this extensive environmental knowledge, otherwise it will die as soon as the researchers leave it to itself.

Thus, it is not so much the fact that life has been created which is highly unlikely, but rather that it has succeeded in adapting to the environment. The emergence of life is perhaps not as extremely rare as Jacques Monod (1972) suspected, says Popper, and he considers the possibility

> that the attempt has been made many, many times - that organisms have arisen, even with some adaptation but not enough adaptation, until one organism survived in the end that was sufficiently well adapted to the environment in which it arose. (Popper 1986/2014, 123)

Sufficient adaptation is extremely unlikely because, as Popper's thought experiment shows, it requires an extraordinary amount of knowledge about the environment.

In this way Popper shows that adaptation is identical to *acquiring knowledge*. What since Darwin has been called 'adaptation' is no longer a final biological explanatory principle, even if it is still in use today. Adaptation itself can now be explained concretely as knowledge acquisition.

Before I go into the knowledge of organisms and the knowledge of cells (in sect. 3.5), I should demonstrate how Popper invalidates the obvious objection of introducing illicit *anthropomorphic expressions* into biology.

The question is: can expressions such as 'pursuing goals', 'having expectations', 'acquiring knowledge', which have a clear meaning in the human world, can such expressions legitimately be transferred to organisms (like plants, animals, cells), organs (like hearts and kidneys), and even to organelles (like mitochondria and ribosomes)?

It is not only the suspicion of anthropomorphism which disturbs the adequate understanding of Popper's active Darwinism. Equally confusing is the fact that explanations in which organisms, organs, and organelles 'pursue goals' or 'have expectations' or 'acquire knowledge' dissolve into virtually nothing when the processes behind these explanations have been reduced to physics and chemistry. Therefore, I will show in sects. 3.3 and 3.4 how Popper solved the problems of reducing biology to biochemistry and biochemistry to chemistry and chemistry to physics.

Only when both the problems of anthropomorphism and reduction are solved can we speak rationally about *knowledge of organisms* and *knowledge of cells* (in sect. 3.5 and 3.6).

3.2 There Is no Biology Without Anthropomorphisms

Thus, the first question is whether the term 'knowledge' is not too much of an 'anthropomorphism'. Can an organism or an organ like the kidney really *know* something?

Popper (1986/2014, 123) says: "I assert that we cannot do biology at all without a certain amount of anthropomorphism". We should treat anthropomorphisms as *theories* that can be right or wrong. For example, the term 'natural selection' is wrong if we understand it anthropomorphically by believing that in biological evolution nature has taken the place of the breeder. Darwin did not believe that. He knew that a breeder could breed black cats with white paws, but natural selection had no such aims. Thus, 'natural selection' is an illicit anthropomorphism because it suggests a false theory.

On the other hand, one can speak of the fact that a dog has a nose, even if it does not look very much like the human nose. Behind this anthropomorphism is the correct theory that the noses of dog and man are 'homologous', i.e. they perform very similar functions. In cases like this, anthropomorphisms must not be banned, otherwise important theories would be excluded from science without further discussion.

Popper's relative tolerance of anthropomorphisms has met with little approval in the scientific community. There is an old dispute in biology about whether organisms or even cell-organelles pursue goals (Montefiore and Noble 1989). It seems that they do because the thesis that scientific explanations succeed better if we talk of goals is testable, for example when we explain the kidneys with the *goal* of purifying the blood. On the whole, the evolution of a kidney is easier to understand when we talk about goals, because only then can we understand problems that had to be overcome on the way to develop this perfectly functioning organ.

Nevertheless, many biochemists shy away from going as far as Popper. They would perhaps speak of goals with regard to animals, but hardly with regard to biochemical processes such as plant photosynthesis, because they rightly presume that all biochemistry is based on chemical reactions. "The fear of using teleological terms", writes Popper (1986/2014, 124), "reminds me of the Victorian fear of speaking about sex". From the great evolutionary biologist J.B.S. Haldane—whom Popper met at a conference of biologists in Hunstanton, Norfolk, in 1936 (Niemann 2014a, chap. I, sect. 6)—the bon mot is handed down: "Teleology is like a mistress to a biologist: he cannot live without her, but he's unwilling to be seen with her in public" (Mayr 1974).

Only many years after Popper's death, a certain circle of biologists did give up this shyness.[9]

3.3 Non-reducibility of Biochemistry to Physics and Chemistry

The problem of teleology is closely related to the scientifically important question of reducibility of biology to chemistry and physics; for if one can describe how amoebae move automatically with the help of special sensors for food in the direction of places with plenty of food, then there will be no need to say that the amoeba pursues the goal of finding food. The higher function of 'pursuing goals' would have been successfully reduced to a biochemical mechanism.

A completely different attitude prevails outside research. In general, people shy away from reducing living things such as humans, animals or plants to physics and chemistry. Henri Bergson expressed what many people think in one way or another: All living things are based on a special, irreducible life force, the élan vital.[10]

However, materialists and Marxists had a different opinion. Like the chemistry of life, called *bio*chemistry, they tried to describe all life processes as pure chemistry. The result was quite successful: Heredity is based on the doubling of a dead molecule, namely DNA, which can be produced synthetically.[11] Photosynthesis is based on the reaction of photons of sunlight with chemical substances. All animal's breathing is based on a complex chemical production cycle resulting in energy packages called 'ATP' (adenosine triphosphate), whose structure was characterized in the 1960s by one of Popper's admirers, the biochemist and Nobel Prize winner Peter Mitchell. These examples are wonderful results of reduction and fully respected science. If only one thing had not gone wrong: All those explanations explained life away.

Do Neo-Darwinists have to rethink? For many of them it had been difficult to change from a Christian or a vitalist's position to Darwin's materialistic view. Now they remained true to their new faith when in 1986, in his great speech to the Royal Society, to biologists and biochemists, to five Nobel Prize winners,[12] Popper wanted to encourage them to rethink Darwinism once again. He asked them not to close their eyes to teleology and irreducibility. However, he was far ahead of his time. In the subsequent discussion Popper and the biochemist and Nobel Prize winner Max

---

[9] Noble (2014b) and others in note 7 above.

[10] In *L'évolution créatrice* (1907), Henri Bergson marks the peculiarity of biological processes with the concept of *élan vital* (French for 'vitality').

[11] Only the *isolated* DNA may be called a dead molecule. As Barbara McClintock (1984) stressed, the genome as a whole is living matter which controls many of its own processes.

[12] Popper (1986/2014). The five Nobel Prize winners in Popper's audience were Hermann Bondi, Peter Medawar, Peter Mitchell, Max Perutz, and George Porter.

Perutz got into a fight about the question of ‘goals or no goals’ which continued smouldering long after Popper’s death. I reported on this in detail elsewhere.[13]

Apparently, there were fears that Bergson’s theory of vitalism would be revived. But Popper’s plea for life was completely different from the views of Bergson or other vitalists, and far away from those of the creationists. The greatest philosopher of science of the twentieth century[14] put forward two theses:

(1) All life is based on biochemistry. (2) Biochemistry cannot be reduced to chemistry and physics. The first thesis is undisputed, the second is not. However, this latter thesis is supported by the purely epistemological and empirically testable argument already mentioned in sect. 3.2: If theories that contain goals, purposes or functions provide a *better explanation* than theories that reduce the things in question to molecular and atomic processes, then the one with the better explanation should be preferred. For the aim of sciences is to find the best of all alternative explanations (Popper 1983), the one that is simpler than all the others and makes more precise and more testable predictions, just to mention the main criteria (Popper 1963/1989, chap. 10, sect. 3, X).

3.4 ‘Based On’ Is an Entirely Different Story Than ‘Explained By’

The reduction of biology to chemistry and physics is in general a desirable objective. Everywhere in scientific research, the aim is to find simpler and more comprehensive laws. However, some reductions do not succeed (Popper 1972), and if they do, they do not necessarily provide a better explanation. For example, the effects of a tornado, such as uprooting trees or tearing off roofs, can be more easily calculated and explained by *macro-effects* such as wind speed and vortex formation than by the micro-events of myriads of air and water molecules on whose dynamics the tornado is *undoubtedly based*. Not even the best computers in the world would be able to calculate the development of concrete storms if they based their calculations on molecular movements. And even if it were possible no one would call such an extremely complex calculation a satisfactory explanation.

Popper’s argument is even better understood when one realizes that ‘based on’ is something completely different than ‘explained by’ (Niemann 2014a, sect. 14). The tornado is *based on* the micro-states of countless air and water molecules, but the macro-effects are better *explained by* the laws of fluidics.

In this sense, it is unsatisfactory to reduce biochemistry or molecular and cell biology to chemistry and physics and leave it at that. All biochemistry is undoubtedly *based on* chemistry and physics; but that *does not explain* everything. The formulas lack something important that is added as soon as biochemists start *to explain* cell

---

[13] Perutz versus Popper, in Niemann (2014a), chap. II, sect. 13.

[14] Peter Medawar, BBC Radio 3, 3 July 1972.

processes in the usual way by talking of the goals or purposes that the many organelles of the cell pursue in a complex interplay.

If Popper is right that "the irreducibility of biology to non-biological sciences, which various people have asserted at various stages, reduces to the irreducibility of biochemistry to chemistry" (Popper 1986/2014, 126), then the old and difficult discussion of whether human beings are nothing more than physico-chemical machines can be answered. This problem boils down to the question of whether our thinking and feeling is nothing more than the activity of neurons. One can now refrain from the 'based on' aspect of highly complex events inside our brain and stick instead to the 'explained by' method. If we succeed in explaining simple physiological processes with the means of biochemistry, we need not go further and reduce them to chemistry and physics on which they are undoubtedly based.

3.5 The Objective Knowledge of Cells, Plants and Animals

Familiar with the irreducibility of biochemistry and the well-considered use of anthropomorphisms, we may now return to Popper's bold assertion that adaptation is closely associated with acquiring knowledge (Popper 1986/2014 (3), 120, 122–123; and Popper 1989c).

> The strange thing is that teleology enters the world with adaptation. Organisms are problem-solvers. Organisms seek better conditions. All of these are thoroughly teleological terms. 'Better conditions' introduce evaluation. And no doubt organisms value, organisms prefer, organisms like this or that better than something else. We cannot avoid all teleological terms, and we cannot avoid all anthropomorphic terms. (Popper 1986/2014, 124)

The talk of knowledge in a cell or in an organelle of the cell such as the chloroplasts that 'know' how to convert air, light and water into sugar and other high-energy hydrocarbons, is also a permissible anthropomorphism, because without it we cannot properly understand biology:

> And how can we avoid applying something like knowledge to animals and, of course, to plants? How can we avoid speaking of the roots of a tree *seeking* for food, *seeking* water, *seeking* better conditions? If we avoid it, then we deceive ourselves and speak in an artificial language when there is no need to introduce an artificial language. (Ibid.)

3.6 Subjective and Objective Knowledge

Thus, it is justified to say that not only a dog has knowledge but also plants and bacteria. But is it the same kind of knowledge we humans have? We must not think that cells know something *consciously* as humans do, when they feel that they know something. The cell's knowledge is not subjective knowledge, and Popper suggests a very strict distinction between *subjective* and *objective* knowledge.[15] Cells contain objective knowledge in the way a book contains knowledge, and it is called objective because it is independent of a knowing subject.[16] Whereas subjective knowledge, as far as we know, is always the product of a working brain.

As for objective knowledge, the cell contains two completely different types. On the one hand, there is digitized knowledge in its DNA. On the other hand, it contains knowledge which is built into its many organelles. To explain this 'built-in knowledge' as opposed to digitized knowledge, imagine a kitchen and a cookbook. Looking around in the kitchen, you will know after a while what the cooker, pots, spoons and so on are for. You can also open the cookbook and read the recipes. In the first case, you decipher the built-in knowledge if your common sense helps you to understand the kitchen equipment. In the second case, you translate the many characters in the cookbook if you know the translation rules. These are the two types of objective knowledge: the *built-in* knowledge and the *digital* knowledge.

The built-in knowledge may be called *analogous* knowledge (Noble 2008b, sect. 3). Because of the logical considerations to be discussed in sect. 6.3, I prefer the term 'built-in knowledge'. Both types of knowledge are objective in the sense of being independent of knowing subjects and they are not subjective knowledge.

3.7 Popper's Evolutionary Epistemology

As for the evolution of organic knowledge there is a striking parallel between Popper's interpretation of human's acquisition of knowledge and Darwin's interpretation of evolution: According to Darwin's theory the individuals who make mistakes have to die; according to Popper the wrong theories have to die. Darwin suggests that all living beings fight a fierce struggle for life; Popper suggests that all theories have to deal with severe criticism. This parallel between evolutionary processes and Popper's falsification method is obvious. However, it is not formulated in the German edition of *Logik der Forschung* from 1934,[17] but only in the English translation *The Logic of Scientific Discovery* twenty-five years later. Since the mid-1950s Popper repeatedly chose Darwinian formulations for his new epistemology such as: "We choose the theory which best holds its own in competition with other theories; the one which, by natural selection, proves itself the fittest to survive" (Popper 1934/1968, sect. 30,

---

[15] The differentiation between 'subjective' and 'objective' in many fields of thinking is a guiding principle Popper has followed since his early years: Popper (1974), chap. 13, Niemann (2012a), sect. 3.

[16] Popper (1972/1979), chap. 3; Popper (1990), 32–39; Niemann (2014a), chap. III, sect. 17.

[17] Compare the formulations of 'Wettbewerb—competition' (sect. 6, p. 19), 'strengen Prüfungen standhalten—withstanding severe proofs' (sect. 30, p. 85), and 'Bewährung—corroboration' (chap. X, p. 237) in Popper's *Logik der Forschung* (1934/2005) with his Darwinian formulations of 1959: "the fiercest struggle for survival" (sect. 6, p. 42); "natural selection" and "the fittest to survive" (sect. 30, p. 108); and "its fitness to survive" (chap. X, p. 251) in the English edition Popper (1934/1968).

p. 108). It was the time when he felt ready to improve Darwin's evolutionary thoughts in his first Spencer Lecture (Popper 1961, sect. 1), as I reported in chap. 2.

It was Donald T. Campbell who coined the expression 'evolutionary epistemology' for the obvious parallel between Darwin's natural selection and Popper's critical epistemology. As we know from their correspondence, both used the expression 'evolutionary epistemology' since 1963 (Campbell 1963, 1974). Popper's 'Sketch of an Evolutionary Epistemology' is a contribution to a seminar in 1970.[18]

These dates of 1963 and 1970 are important, because in the 1970s Konrad Lorenz, Rupert Riedel, and Gerhard Vollmer also presented versions of an 'Evolutionary Epistemology' with capitalized terms. Unlike Popper, these authors explained human and animal cognitive abilities by means of biological evolution referring to the famous example of Konrad Lorenz: jumping from branch to branch without grasping reality properly, an ape would very quickly become a dead ape. In contrast, Popper interprets the cognitive abilities of all living beings as the application of the trial-and-error elimination method which is the only method for acquiring knowledge. And it was Popper who came up with a much more comprehensive theory that generally explains why, *for purely logical reasons*, no cognitive system can acquire new knowledge about the world in any other way than through the logic of trial-and-error elimination[19]:

> It is *logic*, the application of logic to the situation of knowledge (situational logic), that teaches us that knowledge can *only* work with the method of *trial and error*. Thus, the so-called 'evolutionary epistemology' is only an application of logic. In other words, evolution could not proceed otherwise. (Popper 1987)

3.8 Adaptation Is Not the Ultimate Explanatory Principle of Biology

The biological consequences of Popper's evolutionary epistemology are far from being finally explored, considering that it applies not only to the cognitive abilities of humans and animals, but also to plants, bacteria, archaea and even to the still unknown forms of early life which are, for the first time, learning something new about their environment. It is logically impossible that new knowledge passes directly from the environment into the cell or DNA without the method of trial and error. The reasons are the same as those Popper (1972/1979) gave for the impossibility of induction, i.e. deducing theories from experiences. Besides, Popper's evolutionary epistemology gives logical support to the so called 'central dogma of molecular biology' (Niemann 2014a, sect. 19).

According to Popper's logic of knowledge acquisition, adaptation is no longer the ultimate explanatory principle of biology. The question 'Where did some butterflies get their eyespots from?' is no longer answered with 'by adaptation due to specific selection pressure'. Now you can ask a more specific question: 'How do animals,

---

[18] Popper (1972/1979), chap. 2, sect. 16; see also Popper (1986/2015).

[19] Popper (1934/1968), sect. 6 and 30, and Popper (1972/1979), chaps. 1 and 2.

plants, protozoa and all living beings manage to acquire vital knowledge about their environment and above all about their own internal cell life processes?' This is, of course, a question about *objective* knowledge, not about subjective knowledge. For logical reasons discussed in Popper's *The Logic of Scientific Discovery*, the acquisition of this new knowledge is *only* possible by means of the question-and-answer game: by the method of trial-and-error elimination. If one wants to explain life, one has to find out how the first cells and their predecessors realized this logical question-and-answer game in a biological way.

Despite all these novelties, Popper's version of evolutionary epistemology is in harmony with Darwin's ideas of variation and selection. However, Popper suggests a new interpretation; and as Henri Bergson remarked in his *Creative Evolution,* the most thorough progress in science comes often when its results are placed in a new framework. Such a new framework is given by Popper:

- It is not *natural selection* that constitutes the creative element of evolution, but the organisms' search of a better world.
- Darwin's *variation*, for a long time seen as the result of blind mutation, can now be interpreted as proposing alternative theories on adequate knowledge about the environment or the internal chemistry of the cell.
- Darwin's *variation and selection* are not typically biological characteristics, because every kind of newly acquired knowledge, be it in science or in organisms, can, for purely logical reasons, only be obtained by variation and selection, i.e. trial and error elimination.
- Darwin's *selection*, the elimination of the ill-adapted, means now the elimination of false knowledge about the environment. The *lack of knowledge* is the real reason why, evolutionarily speaking, less adapted organisms have disadvantages and are therefore superseded by others.
- Darwin's *adaptation* is no longer an *ultimate* biological explanatory principle, because adaptation can be further explained by knowledge acquisition.[20]

Popper's above mentioned *Logik der Forschung* of 1934 (*The Logic of Scientific Discovery*) has subsequently proven to be an important contribution to biology, because nothing is possible in biology that is not possible in logic. This is of great importance for organisms acquiring new knowledge.

4 A View Into the Innermost Part of Life

4.1 The Distinction Between Information and Knowledge

Popper's logic of research, which includes the logic of biology, is particularly relevant to the question of how first life came into being, because even the earliest precursors

---

[20] Popper (1961), at the end of sect. 2. Popper (1986/2014, 120) writes: "Adaptation is, I suggest, essentially a trial and error learning process that extends over many generations".

of life could not help but gain knowledge about themselves and their environment by the method of trial and error elimination. To do this, they must have developed an appropriate apparatus being able to store and pass on the newly acquired knowledge, initially as analogue or built-in knowledge and later as storable and transferable digital information.

Popper himself did not sufficiently point out the difference between knowledge and information in biology (cmp. Niemann 2014a, sect. 20) because it was self-evident for him that his logic of research and his evolutionary epistemology were not about the transfer of information, but about the acquisition of *new knowledge*.

However, when we read books, we acquire *old* knowledge and do not need to apply the technique of trial and error correction: this kind of getting knowledge is done by *information* transfer. The same applies to bacteria when they exchange digitized information from parts of their DNA on occasion of the so-called ‘horizontal gene transfer’. And it also applies to heredity: doubling and dividing DNA means passing on *information*, rather than acquiring *new knowledge*. The same sort of *information transfer* happens as well when the cytoplasm’s *built-in knowledge* is duplicated during cell division and passed on to the daughter cell. In this biological context, the helpful distinction between ‘knowledge and information’ refers always to the distinction between ‘newly acquired knowledge’ and already stored ‘information’, stored in digitized or analogue form.

From the early beginnings of all life, biological evolution must always have been based on both processes: on the acquiring of new knowledge by learning from trial and error, and on exchanging information that other organisms had previously gained by trial and error elimination.

4.2 All Life Begins with Activity and Knowledge Acquisition

Organisms do not develop formulated theories, but they have *expectations*—objective, not subjective ones. All cells expect that there is air and water and minerals in the world. Mice expect a world where they can hide in the ground. Man expect people who can understand words and sentences. Every organism has built-in expectations, and expectations are like theories.[21] Since the logic of Popper’s *Logic of Scientific Discovery* (1934, 1959) allows only one method for the acquisition of new knowledge, namely inventing alternative theories or expectations followed by reality testing, another element emerges that also sheds new light on evolution and the origin of all life: *activity*. Activity is needed firstly to produce as many alternative theories as possible and secondly to choose the one that best fits reality, or to let the environment choose the best theory.

---

[21] Popper (1994), chap. 8, sect. 1, p. 156. Examples of Popper using theories, conjectures, and expectations as synonyms: in Popper (1963/1989), chap. 1, sect. V, pp. 47–48, p. 71; Popper (1977b), sect. 39, pp. 132f.

Biological activity, unlike volcanic activity, is always focused on a goal, such as finding food or water or a hiding place. Activity is movement with an aim. I think that we have to attribute some sort of activity to even the lowest organisms—to our lowest forefather, or foremother if you prefer. There cannot be adaptation without any aim. There cannot be knowledge without any aim.[22]

Anyone interested in the first beginnings of life or in the transition from inorganic to organic matter, must consider—for logical reasons that can be read about in Popper's writings (Popper 1934/1968; 1963/1989, chap. 10)—how both *activity and knowledge* have been realized even in the simplest forms of life, for without activity and knowledge no life is possible (Niemann 2014a, b, 2013b).

4.3 We Inherit Not Only the DNA, but Above All the Cell

One of Popper's most fertile biological ideas is a combination of obvious triviality and unexpected consequences, as they often occur with him: We not only inherit the genes; we also inherit the cell: "It is the cell that divides, not only the DNA" (1989). "Heredity is not a special function of the genome… The directing staff of a factory must also be duplicated in a daughter factory" (Popper 1989a, 135; 1992a, 1991b). Apparently, Popper means by 'cell' the cytoplasm, the whole cell without DNA, and in the following I will use the word 'cell' in this sense.

If one thinks a little further, a fascinating aspect quickly becomes apparent: Human *genes* are strung together in 46 dead DNA molecules; on the other hand, the *cell* consists of trillions of molecules forming cytoplasm which includes mitochondria, ribosomes, membranes and other organelles, and whatever else maintains life. We inherit all our cells only from our mother, because all our body cells specialized in certain tasks are copies of the fertilized egg cell. Each mother inherited the egg cell with the built-in machinery of life from her mother. The actual machinery of life, which consists of knowledge and activity and which can therefore make sense of the DNA as a *recipe book* for the production of various proteins, is passed on via the maternal line only. Each human being is therefore related to his two grandmothers in very different ways. We inherit a quarter of our genes from each grandmother; but only from our maternal grandmother do we inherit the cellular apparatus and therefore the entire machinery of life, which includes the activity of the cell and the knowledge of the cell, for example the knowledge of how DNA-information is brought to life (Niemann 2014a, 98–100).

We must therefore distinguish between analogue and digital inheritance in the same sense as we distinguish between analogue and digital knowledge.

---

[22] Popper (1986/2014), 125; cmp. 'agency' in Noble and Noble (2018).

4.4 Looking at Genes the Wrong Way

“We are, it seems to me, looking at genes the wrong way” wrote Popper (1989a, 135) to a South African friend and biochemist. He discussed the contents of this letter at length with Günter Wächtershäuser, and he also sent copies to Peter Mitchell and William Bartley. Then, from the ideas outlined in this 1989 letter, he drew up two lectures given at the University of Santander in 1991 and 1992 with the meaningful titles “Putting Genetics in Its Place” (Popper 1991b) and “An Enzymatic Theory of Self-Correcting Evolution” (Popper 1992a). Popper tried to show that the genome does not play a leading role, neither in the cell and nor in evolution.

Earlier than Popper, and defying the spirit of the times as well, Barbara McClintock (1984), winner of the Nobel Prize 1983, had the same view: “The genome is not the dictator of cells”. The topicality of Popper’s advance is not diminished by his forerunner; for even today far too few biologists, not to mention philosophers, recognize the significance of the new biology. However, evolutionary biologists like Eva Jablonka (1999) and physiologists like Denis Noble understand and spread it as a revolution: “Physiology is rocking the foundations of evolutionary biology” (Noble 2013). This physiology that shakes the foundations of evolutionary biology stands for the new view on physiological processes inside cells and organisms. Like Popper, but equipped with empirical evidence, Denis Noble argues against the gene-centred view of natural selection and the so called ‘central dogma of molecular biology’ (Noble 2008a, 2014b). He was surprised when, in 2014, he discovered in *Karl Popper and the Two New Secrets of Life* (Niemann 2014a) the preparatory work Karl Popper had done almost thirty years earlier, and summed up:

> He [Popper] proposed a radical interpretation of Darwinism, essentially rejecting the Modern Synthesis, by proposing that organisms themselves are the source of the creative processes in evolution, not random mutations in DNA. Popper suggested Darwinism was not so much wrong, but seriously incomplete. He also stated that biochemistry (and so *a fortiori* physiology) could not be reduced to physics and chemistry. Many of the points made in the recent special issue of *J. Physiol.* (Noble et al. 2014b) *were* therefore made nearly 30 years ago (Noble 2014a).

5 Evolution as an Adventure of the Mind

Theories of creation and evolution have always strongly influenced our world view. Darwinism, with its central concept of the “struggle for life” (in the title of Darwin 1959), inspired the powerful political ideology of Social Darwinism. George Bernard Shaw alluded to this Social Darwinism when he wrote in retrospect of World War I: “Neo-Darwinism in politics had produced a European catastrophe of a magnitude so appalling, and a scope so unpredictable, that as I write these lines in 1920, it is still far from certain whether our civilization will survive it.” (Shaw 1921/1987, introduction).

As we all know, the worst was yet to come: the doctrine of the presumptuous Aryan superhumans and the alleged Jewish sub-humans of the following decades was also based on false interpretations of Darwin's theory.

Popper's improvements to Darwinism have the potential to establish a completely different tradition in which another feature of nature becomes a model for us today: In ancient times, matter had once managed to collect objective knowledge about itself—unconscious knowledge of course—and to constantly increase and improve this knowledge by reality checks. In early organisms, knowledge initially grew up as built-in knowledge or analogous knowledge (Popper 1990, 32–39). Over time, the unicellular organisms from which we all ultimately descend started storing knowledge also in digital form. Only then a lively exchange of experiences became possible between the world's first libraries called DNA and RNA. In this way knowledge grew and grew and was saved in DNA, RNA, and cytoplasm, later also in brains and in traditions; eventually it was discussed and improved by means of language and stored in books and libraries for later rediscovery and further processing.

In one of these libraries, in the 'Karl Popper Sammlung' at the University of Klagenfurt, I found a handwritten note by Karl Popper in which he summarized four billion years of organic life in a wonderfully memorable statement that could be the guiding idea for our future beyond Social Darwinism: "The whole evolution is an adventure of the mind" (Popper 1986, sect. VI, p. 20).

6 Conclusions and Consequences

6.1 From Biological to Cosmological Evolution

In this paper I have tried to outline some of Popper's thoughts on evolutionary biology; for further reading I have added the relevant literature references. Elaborated in detail, Popper's evolutionary ideas could easily fill a whole book. It was at the age of twelve that he began to think about evolution, and his biological thinking only came to an end in July 1994, two months before his death, when he gave his last interview on how thoughts can become neuronal impulses (Popper et al. 2010). His most important achievement is, in my opinion, his concept of 'world 3' as man's evolutionary exosomatic tool. Just as spiders have their webs, birds have their nests, beavers have their dams, so man has words, his imagined, spoken or written words, and with these words a whole world of problems and theories emerges that enabled the universe to reflect on itself, and ultimately, in the form of science, changed the earth to an extent comparable to the work of the organisms that created the atmospheric oxygen shell or the coral reefs or the limestone Alps (Popper 1972/1979, introduction of chap. 8).

Popper's biological ideas can lead us to a better understanding not only of the biological but also of the cosmological evolution (Niemann 2017). Strangely enough it seems that Popper's concept of built-in knowledge is transferable into the realm of inorganic chemistry, the world of atoms and molecules. Apparently, much more knowledge can be read out from water (H_2O) and all its physical and chemical properties than from the two atoms H and O on which it is based.

6.2 From Amoeba to Einstein

In my opinion, this 'reading out' of objective, non-human knowledge is one of the most interesting starting points for expanding Popper's evolutionary biology into an evolutionary cosmology in which the entire evolution is indeed "an adventure of the mind". To do this I have to start—how else could it work?—with criticism of Popper's dictum that there is only one essential step from amoeba to Einstein, namely the step of letting one's theories die in place of oneself: "The main difference between Einstein and an amoeba… is that Einstein *consciously seeks for error elimination*", while the amoeba dies when it makes mistakes.[23]

Popper knew, of course, that it was not really just one step that led evolution to science. He certainly would have agreed that at least seven steps are biologically relevant and philosophically interesting, profound and further-reaching: (*i*) The first built-in knowledge. (*ii*) The first theoretical knowledge in the form of non-verbal expectations and goals of single-celled organisms and plants. (*iii*) Digitalized DNA/RNA knowledge. (*iv*) Brain stored knowledge. (*v*) Verbalization of built-in or imagined expectations and descriptions. (*vi*) A level of consciousness making it possible to talk about one's own mental products in order to be able to apply self-criticism. (*vii*) The world of written words which may lead to discussing problems with others in order to let false theories die, especially in the context of science.

6.3 Human Knowledge Verses Biological Knowledge

Of these seven problem areas, I will confine myself only to the first one: built-in knowledge. As already said, evolution can, for entirely logical reasons acquire new knowledge only by trial and error.

Popper explained this already in his *Logik der Forschung* in 1934. Three years later he published his tetradic scheme, also called *Popper Scheme*, his very own

---

[23] Popper (1972/1979), chap. 1, sect. 10, pp. 24–25; chap. 2, sect. 16, p. 70; and Appendix I, sect. 4, p. 347.

contribution to this universal epistemology relevant not only for science but also for ethics and metaphysics[24]:

$$P_1 \rightarrow TT \rightarrow EE \rightarrow P_2$$

P_1 = initial problem situation
TT = tentative theories (problem solutions)
EE = error elimination and selection of the best solution
P_2 = new and improved problem situation

The logic of Popper's epistemology is similar to the logic of evolution. In order to ask questions and find experimental solutions, even the simplest creatures need activity. Where this activity comes from is still one of the unsolved mysteries of life. Martin Heisenberg and Björn Brembs used Drosophila flies to demonstrate that this type of problem-solving activity does exist and that even simple organisms are problem solvers,[25] in this respect as ingenious as apparently all living creatures from blue-green algae to humans.

What these brilliant creatures gain by repeatedly using the $P_1 \rightarrow P_2$ scheme is knowledge about their environment and, especially in pre-cellular times, knowledge about their own internal biochemical world. This knowledge is, of course, objective, not subjective, as discussed in chap. 3 above. However, is Popper right to call this built-in knowledge 'objective knowledge'?

I propose to differentiate between the objective built-in knowledge of unicellular organisms, plants, or animals, and the human knowledge of man and science.

Firstly: Only in human knowledge, especially in written knowledge, do logical relations exist, such as contradictions, equivalences and deductive consequences. These relations exist whether someone has discovered them or not. But they do not exist in built-in knowledge: factual consequences are not the same as logical consequences; and factual opposites, such as 'up—down', may look like contradictions but they are not logical contradictions. Built-in knowledge has no *logical* consequences and parts of it cannot *logically* contradict other parts. Logical relations only arise when a person analyses the built-in knowledge and expresses it in language; for logic is a product of language and does only exist in the world of language, a world that Popper (1977b) called 'world 3'.

Secondly: Because biological knowledge belongs to the real world 1 rather than to world 3, it is not fallible (in a logical sense) and therefore has a completely different character than human theories or conjectures or expectations. This fact is very much at odds with one of Popper's best proven theories: the infallibility of *all* knowledge. However, it is consistent with Popper's theory of fallibilism and falsification which

---

[24] Briefly in Popper (1934/1968), more detailed in: 'What is dialectic?' (1937), Popper (1963/1989), chap. 15, sect. 1, first three paragraphs; also in Popper (1972/1979), chap. 8, 287–288. Applied to ethics in: Niemann (1993).

[25] Heisenberg (2009), Brembs (2008). Bacteria are also problem solvers, e.g. when it comes to attacking an organism together: Waters and Bassler (2005).

only applies to *verbally* formulated theories that can only be falsified by *verbal* statements.

Nevertheless, biological knowledge is no metaphor, but real and objective knowledge, because it can be ‘read out’ from the cell or an organelle. Not ‘read out’ as from a digital memory stick or DNA, because for this ‘reading out’ of biological knowledge, for this translation of world 1 objects into the world 3 of language, a mediator is needed. This mediator is consciousness or Popper’s ‘world 2’ which plays here the same important role as in his well-known *interaction theory*, in which World 2 is indispensable when it comes to transforming the world 1 printer ink into thoughts that belong to both the mental world 2 and the world 3 of descriptions, problems or theories (Popper 1977b). Consciousness or ‘world 2’ is also needed as a mediator when the transformation goes in the opposite direction, when objects of world 3, such as written books or printed plans, cause huge changes in the physical world 1. One example is construction plans for a dam, which can only change the landscape if man’s world 2 intervenes.

The ‘reading out’ of biological knowledge from a biological cell works in a similar way as when an engineer reads out the knowledge built into a camera that he has never heard of and knows nothing about. He takes it apart and studies its functions with growing understanding. In the end, he knows something he has obviously learned from the camera. Although this knowledge was gained by the intervention of the engineer’s subjective world 2, it has now got a new status: It has become objective knowledge about a part of the biological world 1. Separated from the cell, it has become an object of the engineer and only now it is criticizable and improvable. Now logic is applicable, and the engineer can deduce, for example, from scientific findings that he should change a certain gene in order to improve the metabolism of the cell. The cell cannot do such a thing. It cannot but apply the trial-and-error method of natural selection, in order to improve its knowledge about its environment or internal biochemistry.

6.4 Verification: Learning from Confirmations, not from Mistakes

But even without the capacity for logic and criticism the amoeba’s knowledge is real knowledge about its environment and the biochemistry of its internal life, a knowledge that has constantly been growing in the course of evolution and now—when it is read out—has reached the dimension of a large library. It is this huge knowledge that gives cellular life a permanence over generations far exceeding the stability of the atmosphere, oceans, continents and mountains, for these have all changed tremendously, while bacteria and the basic configuration of the cell in organisms have remained almost the same for over three billion years. And in the meantime, the built-in ‘theories’ of the cell (theories, if we read them out) have lived incomparably longer than any man-made theory.

This chain of learning consists of many links, each of which did withstand natural selection in its time and was saved from extinction. As much as natural selection resembles Popper's theory of learning if we interpret extinction as 'falsification', and as much as this analogy is the point of every 'Evolutionary Epistemology' (see "chap. 3.7" above), there is another biological learning process that is at least as important as Popper's or Darwin's error elimination, a learning process which Popper would not have liked at all: *verification.*

To explain *verification* as a method of biological discovery as briefly as possible: If an inventor of a new engine tries out ten ideas at the same time and fails, he will not learn the cause of his failure. He would waste time and probably also his whole life by trying out ten ideas at once every week.[26] In order to learn from mistakes, it would have been better to proceed step by step and check each part individually, i.e. with the Popper method of falsification. However, it would have taken him and his successors many thousands of years, even if they had tested one possible combination every day. Nevertheless, it is not logically impossible that he would have been lucky after only a short time if all ten ideas had fitted together at the first go and a new engine with enormous efficiency had been created. In this case, our lucky engineer would have been able to *verify* ten ideas at once and *prove* that he had found a feasible option right away.

To put it in the language of evolution: a DNA that has varied ten cell properties in one go, say ten enzymes for digesting a new type of food, DNA enabled its carrier to discover a previously uninhabited new habitat or 'possibility space' as Popper (1990, 19) called it.

It is this verification method that has allowed evolution to make great leaps in a short time without wasting time on countless falsifications. Such holistic experiments are realized as the well-known processes of *sexual recombination* and *horizontal gene transfer*. Both processes ensure that several genetic innovations (e.g. through mutations) occur in one and the same DNA and can be tested as a package by the environment. The time-consuming step-by-step falsification is avoided by the fact that billions of bacteria are simultaneously tested by the environment with the effect that the 3.6 million permutations of ten properties are tested in no time. It is therefore possible that a new bacterial species makes a huge leap in development, while the others more or less wither away because they cannot keep up with this single specimen in terms of reproduction and viability. When this happens, biological proof or verification has been provided: a new and better habitat has been discovered or an improved cellular inner world has been created, for example with better energy and protein production.

---

[26] Actually, with one test per day he needs additional 10 000 years for testing all 10! permutations.

6.5 Verification Is not a Disguised Falsification

Of course, any verification can be reinterpreted as a falsification, just as a "falsification of a statement *a* can always be interpreted as the verification of its negation *non-a*" (Popper 1983, 181) But in the case of an evolutionary verification, the price of reformulating it into a falsification is that, contrary to the aim of science, the simpler explanation is abandoned in favor of the more complicated one.

The one lucky bacterium among billions of bacteria that has changed ten properties in one go and survived all its relatives is alive because it has survived natural selection, whereas all others have been eliminated. This is in accordance with the teachings of Darwin—but not with the teachings of Popper. According to Popper, the survivor, be it a machine, a computer program, an individual or a theory, took one step at a time and learned from his mistakes. In our biological case, of course, the billions of losers learned nothing from their mistakes, because they all died very quickly. The only one that has learned something is our lucky bacterium, but it is *not from mistakes* that it has learned something. It rather has applied a clever method to discover unoccupied habitats, but this method is not a research procedure in which one systematically learns from mistakes. The irony of the story is that the English title of Popper's *Logic of Scientific Discovery* is a mistranslation: this book contains no logic of scientific discovery—there is no such thing—but a logic of falsification. Of course, there are scientific discovery procedures, good ones and bad ones. But the researcher, who always puts ten variations to the test at once, would probably have been fired before wasting too much money and time.

The heyday of the biological verification method was probably when the machinery of life had to be discovered and developed, the period before the appearance of the three evolutionary kingdoms of archaea, bacteria and prokaryotes.

6.6 Concluding Remark

I could only discuss a few of Popper's evolutionary biological thoughts here.[27] Although they are not yet very far developed, they have high transition probabilities to lead to new 'possibility spaces' (Popper 1990, 17–22; Niemann 2015, 2017), where we can acquire so much knowledge that the transition probabilities continue to grow so that we can enter further fascinating possibility spaces. This constant discovery of unoccupied new possibility spaces or habitats is exactly what has happened during the almost four billion years of evolution—and it is exactly what Popper saw as the main idea of evolution: "All evolution is an adventure of the mind" (Popper 1986, 20).

---

[27] More in my upcoming book on the method of discovery in biological and inorganic evolution.

Acknowledgments 'KPS A:B' stands for 'Karl Popper-Sammlung', the Popper archive of the Klagenfurt University Library, box A, folder B. I thank Thomas Hainscho and the Karl-Popper-Sammlung for permission to quote from the archived material, and Nicole Sager and Lydia Zellacher for supporting my archive work. Popper's estate is also available under the same signatures in the Archives of the Hoover Institute, California.

References

Baldwin, J.M.: A new factor in evolution. Am. Nat. **30**(441–451), 536–553 (1896)

Bartley, W.W.: Philosophy of biology versus philosophy of physics. In: Bartley and Radnitzky, 7–45 (1987)

Bartley, W.W., Radnitzky, G. (eds.): Evolutionary epistemology, rationality, and the sociology of knowledge. Open Court, La Salle, IL (1987)

Brembs, B.: The importance of being active. J. Neurogenet. **23**, 1–7 (2008). https://doi.org/10.1080/01677060802471643

Campbell, D.T.: Letter to Karl Popper, 3 June 1963, 1, unpublished, KPS 282:12

Campbell, D.T.: Evolutionary epistemology. In: Schilpp, P.A. (ed.) The Philosophy of Karl Popper, vol. II, pp. 423–463. Open Court, La Salle, IL (1974)

Darwin, Ch.: On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London (1859)

Heisenberg, M.: Is free will an illusion? Nature **459**, 164–165 (2009)

Jablonka, E.: Epigenetic inheritance and evolution: The Lamarckian dimension. Oxford University Press, Oxford (1999)

Mayr, E.: Teleological and teleonomic: a new analysis. In: Cohen, R.S., Wartofsky, M.W. (eds.) Methodological and Historical Essays in the Natural and Social Sciences, 91–117, Boston Studies in the Philosophy of Science, vol. 14, Reidel, Dordrecht–Boston (1974)

McClintock, B.: The significance of responses of the genome to challenge. Nobel Prize Speech 1983. Science **226**, 792–801 (1984)

Monod, J.: Chance and necessity (French 1971). Translated by A. Wainhouse. HarperCollins, New York (1972)

Montefiore, A., Noble, D. (eds.): Goals, no-goals and own goals. Unwin Hyman, London (1989)

Niemann, H.-J.: Die Strategie der Vernunft. Rationalität in Erkenntnis, Moral und Metaphysik [The strategy of reason. Rationality in epistemology, ethics, and metaphysics]. Vieweg, Braunschweig-Wiesbaden 1993

Niemann, H.-J.: Die Chronologie eines lebenslangen Interesses [chronology of a life-long interest]. In: K.R. Popper, Wissen und das Leib-Seele-Problem [Knowledge and the body-mind problem], 510–546. Mohr Siebeck, Tübingen (2012a)

Niemann, H.-J.: Geist als ein Kraftfeld. Bemerkungen zu Karl Poppers Ideen von 1991 [Mind as a force field. Remarks on Karl Popper's ideas of 1991]. In: Aufklärung und Kritik 4, 13–22 (2012b)

Niemann, H.-J.: Alle Lebewesen steuern ihre eigene Evolution. Bemerkungen zu Karl Poppers Medawar-Vorlesung 1986 [All living beings control their own evolution. Remarks on Karl Popper's Medawar Lecture]. Teil I. Aufklärung und Kritik 1, 21–39 (2013a)

Niemann, H.-J.: Alles Leben ist Wissen und Wollen. Karl Poppers Evolutionsbiologie [All life is knowledge and will. Karl Popper's evolutionary biology]. In: Neck, R. (ed.), Evolution—Natur, Mensch, Gesellschaft, 239–261. Peter Lang, Frankfurt/New York/Vienna (2013b)

Niemann, H.-J.: Karl Popper and the two new secrets of life, including Karl Popper's Medawar Lecture and three related texts. Mohr Siebeck, Tübingen (2014a)

Niemann, H.-J.: Karl Popper und der biologische Ursprung von Aktivität, Wissen und Freiheit [Karl Popper and the biological origin of activity, knowledge, and freedom]. Aufklärung und Kritik 1, 7–22 (2014), quoted as Niemann (2014b)

Niemann, H.-J.: Eine Welt der Möglichkeitsräume [a world of possibility spaces]. In Popper (2015), 471–474, quoted as Niemann (2015)

Niemann, H.-J.: Utopias, Wonders, and a World of Possibility Spaces. A Short Introduction into the Theory of Possibilism. https://doi/org/ 10.13140/RG.2.2.30300.62084. German: H.-J. Niemann, ›Über Wunder und Utopien – Eine Welt voller Möglichkeitsräume‹, in Aufklärung und Kritik 3 (2017), 7-38

Niemann, H.-J.: Karl Poppers Spätwerk und seine 'Welt 3' (Popper's late work and his world 3). In: Franco, G. (ed.), Handbuch Karl Popper. Springer VS, Wiesbaden 2019, 101–118; https://doi.org/10.1007/978-3-658-16239-9_6, quoted as Niemann (2018a)

Niemann, H.-J.: Popper, Darwin und die Biologie. In: Franco, G. (ed.), Handbuch Karl Popper. Springer VS, Wiesbaden (2019), 359–380; https://doi.org/10.1007/978-3-658-16239-9_25, quoted as Niemann (2018b)

Noble, D.: The music of life. Oxford University Press, Oxford (2008)

Noble, D.: Genes and causation. Phil. Trans. R. Soc. A 366, 3001–3015 (2008); https://doi.org/10.1098/rsta.2008.0086, quoted as Noble (2008b)

Noble, D.: Physiology is rocking the foundations of evolutionary biology. Exp. Physiol. **98**(8), 1235–1243 (2013)

Noble, D.: Secrets of life from beyond the grave. Physiology News. December 2014, 34–35 (2014a)

Noble, D., et al.: Evolution evolves: physiology returns to the centre stage. J. Physiol. **592**, 2237–2244 (2014)

Noble, R., Noble, D.: Harnessing stochasticity. Chaos **28** 106309 https://doi.org/10.1063/1.5039668 (2018)

Popper, K.R.: The logic of scientific discovery (German: Logik der Forschung, 1934, Julius Springer, Vienna 1935, 11th edition, Mohr Siebeck, Tübingen 2005). Hutchinson, London 1959; Routledge, London 1992 and later. I quote from the 2nd edition of 1968 as Popper (1934/1968)

Popper, K.R.: The poverty of historicism. Routledge, London (1957)

Popper, K.R.: Evolution and the tree of knowledge. Herbert-Spencer Lecture, Oxford, 30 October 1961. In: Popper (1972/1979), chap. 7, quoted as Popper (1961)

Popper, K.R.: Conjectures and refutations. Routledge and Kegan Paul, London (1963/1989)

Popper, K.R.: Realism versus idealism. Lecture, 6 April 1970. KPS 105:10, quoted as Popper (1970)

Popper, K.R.: Scientific reduction and the essential incompleteness of all science. In: Ayala, F.J., Dobzhansky, T. (eds.) Studies in Philosophy of Biology, 259–284, McMillan, London 1974, quoted as Popper (1972)

Popper, K.R.: Objective knowledge. 5th ed. Clarendon, Oxford 1979, quoted as Popper (1972/1979)

Popper, K.R.: Unended quest: an intellectual autobiography (1974). Fontana, London 1976, quoted as Popper (1974)

Popper, K.R.: Natural selection and the emergence of mind. Darwin-Lecture, Cambridge, 8 Nov 1977. In: Dialectica, Vol. 32, Fasc. 3/4, 339–355 (1978); KPS 203: 1–4; also in Bartley and Radnitzky (1987), 139–155, quoted as Popper (1977a)

Popper, K.R.: The self and its brain. In: Popper, K.R., Eccles, J.C. (eds.) The Self and Its Brain. Part I. Springer, New York (1977), quoted as Popper (1977b)

Popper, K.R.: Realism and the aim of science. Routledge, London/New York (1983)

Popper, K.R.: Eine Weiterentwicklung des Darwinismus [Further development of Darwinism]. Lecture in Vienna, Billrothhaus, 15 March 1986; archived in KPS 493:20; to be published in K.R. Popper, Objektive Erkenntnis, Mohr Siebeck, Tübingen 2021; quoted as Popper (1986)

Popper, K.R.: A new interpretation of Darwinism. The first Medawar lecture 1986; firstly published in German by H.-J. Niemann in: Aufklärung und Kritik 1, 7–20 (2013); English in: Niemann (2014a), Appendix A, 115–128, quoted as Popper (1986/2014)

Popper, K.R.: Die Erkenntnistheoretische Position der Evolutionären Erkenntnistheorie (1986) [The epistemological position of the evolutionary epistemology]. German, in Popper (2015), chap. 15, 382–396, quoted as Popper (1986/2015)

Popper, K.R.: Letter to Hans Albert, 18 Oct 1987. In: Morgenstern, M., Zimmer, R. (eds.), Hans Albert, Karl Popper, Briefwechsel [correspondence Popper-Albert]. Fischer, Frankfurt/M. (2005), p. 261, quotation translated by H.-J. Niemann, quoted as Popper (1987)

Popper, K.R.: Letter to a friend. In: Niemann (2014a), Appendix D, 134–138, quoted as Popper (1989a)

Popper, K.R.: Lecture on molecular biology. European molecular biology laboratory, Heidelberg, 12 April 1989, unpublished, KPS 500:10, quoted as Popper (1989b)

Popper, K.R.: A world of propensities. Thoemmes, Bristol (1990)

Popper, K.R.: Das Geist-Gehirn-Problem—eine versuchsweise Lösung. English original: the mind brain problem. A Conjectured Solution, August 1991, unpublished, KPS 508:3. Posthumous translated und published by H.-J. Niemann. In: Aufklärung und Kritik 4, 7–12 (2012), quoted as Popper (1991a)

Popper, K.R.: Putting genetics at its place. Lecture at University of Santander, July 1991, unpublished, KPS 504:11, quoted as Popper (1991b)

Popper, K.R.: An enzymatic theory of self-correcting evolution. Lecture at University of Santander, 14 July 1992, unpublished, KPS 504:10, quoted as Popper (1992a)

Popper, K.R.: In search of a better world. Routledge, London (1992b)

Popper, K.R.: Towards an Evolutionary Theory of Knowledge (1989). In K.R. Popper, A World of Propensities, Bristol: Thoemmes 1990, part II, quoted as Popper (1989c)

Popper, K.R.: Knowledge and the Body-Mind Problem. Kenan Lecture, Atlanta, April and May 1969. London: Routledge 1994

Popper, K.R.: The myth of the framework. Routledge, London (1994)

Popper, K.R.: Erkenntnis und evolution [knowledge and evolution], in German, edited by H.-J. Niemann, Mohr Siebeck, Tübingen (2015)

Popper, K.R., Lindahl, B.I.B, Århem, P.: The conciousness-brain relation (1994). KPS 548:19; published in German only: Die Beziehung zwischen Bewusstsein und Gehirn: Diskussion einer interaktionistischen Hypothese. Edited by B.I.B. Lindahl, translated by H.-J. Niemann. In: Aufklärung und Kritik 4, 9–21 (2010)

Shaw, G.B.S.: Back to Methuselah (1921). Penguin, London (1987), quoted as Shaw (1921/1987)

Spencer, H.: Principles of biology. Williams and Norgate, Edinburgh-London (1864–1867)

Wagner, A.: Arrival of the fittest: solving evolution's greatest puzzle. Penguin, New York (2014)

Waters, C.M., Bassler, B.L.: Quorum sensing: Cell-to-cell communication in bacteria. Annu. Rev. Cell Dev. Biol. **21**, 319–346 (2005)

Hans-Joachim Niemann Dr. rer. nat (physical chemistry). Former head of a research group at Siemens AG; numerous patents. Freelance writer on problem-solving ethics and critical rationalism. 25 German first editions of Popper's writings; and 20 publications on his philosophy; in English: "Karl Popper and the Two New Secrets of Life" (2014). Editor and translator of six volumes of Collected Papers in German Language (Mohr Siebeck, Tübingen). Email: hjn@h-j-n.de

The Arkansas Creationism Trial Forty Years On

Michael Ruse

1 Prologue

Q: In connection with the attributes of science and this issue of testability, does the concept of falsifiability mean anything to you?

A: Yes. The concept of falsifiability is something which has been talked about a great deal by scientists and others recently. It's an idea which has been made very popular by the Austrian-English philosopher, Karl Popper. Basically, the idea of falsifiability is that there must be, as it were, if something is a genuine scientific theory, then there must, at least, conceivably be some evidence which could count against it. Now, that doesn't mean to say that there's actually going to be evidence. I mean, one's got to distinguish, say, between something being falsifiable and something being actually falsified

But what Popper argues is that if something is a genuine science, then at least in the fault experiment, you ought to be able to think of something which would show that it's wrong.

For example, Popper is deliberately distinguishing science from, say, something like religion. Popper is not running down religion. He's just saying it's not science. For example, you take, say, a religious statement like God is love, there's nothing in the empirical world which would count against this in a believer. I mean, whatever you see—You see, for example, a terrible accident or something like this, and you say, "Well, God is love. It's free will," or, for example, the San Francisco earthquake, you say, "Well, God is love; God is working his purpose out. We don't understand, but nothing is going to make me give this up."

Now, with science, you've got to be prepared to give up.

Q: I was going to ask you for an example of falsifiability in the realm of science.

M. Ruse (✉)
Florida State University, Tallahassee, FL 32306, USA
e-mail: mruse@fsu.edu

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_14

A: Well, let's take evolutionary theory, for example. Suppose, I mean, contemporary thought on evolutionary theory believes that evolution is never going to reverse itself in any significant way. In other words, the dodo, the dinosaurs are gone; they are not going to come back.

Suppose, for example, one found, say, I don't know, somewhere in the desolate north up in Canada, suppose one found evidence in very, very old rocks, say, of mammals and lots and lots of mammals and primates, this sort of thing, and then nothing for what scientists believe to be billions of years, and then suddenly, mammals come back again.

Well, that would obviously be falsifying evidence of evolution theory. Again, I want to make the point, you've got to distinguished between something actually being shown false and something being in principle falsifiable. I mean, the fact that you've got no contrary evidence doesn't mean to say that you don't have a theory. I mean, it could be true.

Q: The last characteristic you mentioned was that science was tentative. Can you explain that characteristic of science?

A: Yes. Again, this is all very much bound up with the points I've been making earlier. What one means when one says that science has got to be tentative is that somewhere at the back of the scientist's mind, he, or increasingly she, has got to be prepared to say at some point, "Well, enough is enough; I've got to give this theory up." It doesn't mean to say you are going to be every Monday morning sort of requestioning your basic principles in science, but it does mean that if something is scientific, at least in principle, you've got to be prepared to give it up.[1]

1. *Background*

In 1957, the Soviet Union launched an artificial satellite that orbited the planet, Sputnik. It was the height of the Cold War and recognized at once as a huge propaganda success for the Russians. Appalled, America set about responding and, in the post-mortems following Sputnik, it became clear to all that American science education, particularly at the school level, was in dreadful shape. Money and resources were poured into organizations formed to improve such education, and in 1958, the American Institute of Biological Sciences founded the Biological Sciences Curriculum Study (BSCS) to tackle issues of high-school biological education. In the US, education is under state (not federal) control, so the strategy taken was that of producing high-quality textbooks, that could then be marketed at knock-down prices, thus attractive to school boards looking for material for classes. Among the offerings was *Biological Sciences: Molecules to Man* (1963). It is very thorough, covering all aspects of biology, including Darwin's theory of natural selection, which is presented as the correct explanation of organic origins, including human origins.

This was a major innovation. In 1925, in Dayton Tennessee, the school-teacher John Thomas Scopes was prosecuted for breaking state law by teaching evolution (Larson 1997). Although convicted, the verdict was overturned on appeal, and that was the end of matters. Except not quite. Scopes may have been found not-guilty but the trial had a chilling effect on American science education. As a matter of policy, text-book publishers go for a lack of controversy. They want to sell their books all over America, including the conservative, evangelical South. Evolution was dropped, and no one got it, north or south. Now, thanks to Sputnik, it was back on the agenda, flauntingly so.

As it happens, a number of southern states had anti-evolution laws, but by the mid-sixties, educators in these states wanted to get on-board with the new direction in science, and so the BSCS books were adopted. The State of Arkansas, which had an anti-evolution law on its books, fought back, and, thanks to counter-resistance by evolutionists, backed by the American Civil Liberties Union (ACLU), the case—*Epperson versus Arkansas*—went to the Supreme Court. The anti-evolution law was struck down as a violation of the First Amendment separation of Church and State. The premise: "The overriding fact is that Arkansas' law selects from the body of knowledge a particular segment which it proscribes for the sole reason that it is deemed to conflict with a particular religious doctrine; that is, with a particular interpretation of the Book of Genesis by a particular religious group." The conclusion: "[T]he state has no legitimate interest in protecting any or all religions from views distasteful to them."[2]

That seemed to be that. But not quite. The biblical literalists, formerly known as "Fundamentalists," now more commonly as "Creationists" (or "Scientific Creationists") fought back. They had a formidable weapon. In 1961, two literalists, John C. Whitcomb, a Princeton-trained biblical scholar, and Henry M. Morris, a hydraulic engineer, co-authored *Genesis Flood: The Biblical Record and Its Scientific Implications.* It became the bible (if one might use such a metaphor) of the literalist movement. Pushing the doctrine of "Young Earth Creationism," the authors claimed that every word of the bible, read literally, is supported by modern science. The focus is on Noah's Flood. Geology shows the Earth is recent, and that at some point it was covered with water; the fossil record shows that evolution is untrue and is more consistent with the pre—and post-Flood biblical accounts of animals; and much, much more along the same lines. Showing that the Cold War was a factor influencing all sides, the reason for the focus on the Flood rather than (say) Adam and Eve and the Garden of Eden, was that, like many evangelicals, the authors were "dispensationalists," believing that history is divide into periods, showing God's revelation and plan of salvation. The Flood marks the end of the second (the first is Adam and Eve being kicked out of Eden) and the reason for its great contemporary importance is that it is a harbinger of Armageddon, which is going to come shortly and end our, final dispensation (Numbers 2006).

Paralleling the *Origin of Species* was *Genesis Flood*! However, after *Epperson versus Arkansas,* the literalists—let us now call them "Creationists"—had to change strategy. Before, the aim could be simply to exclude evolution entirely and to force *Genesis Flood*'s young-earth creationism upon science education. Now, evolution could not be kept out, so the new aim became forcing "balanced treatment" upon science education. If you taught evolution, then you had to teach creationism, in a parallel and non-prejudicial manner. Thus, things went into the 1970s, with the

Creationists seemingly having the upper hand. They were barred obviously from the science journals, so they started inviting evolutionists to debate origins—Darwin versus the Bible. Except the claim was that it wasn't the bible that was being debated. It was "Creation Science," supposedly a perfectly legitimate science that offered a different account of origins from the evolutionists. That it copied word for word the stories of the bible was technically irrelevant.

One should say the Creationists—notably Henry Morris and his partner, Duane T. Gish (who had a conventional biological education doing his doctorate at Berkeley)—had considerable success in the debates, held on campuses (needing masses of space, often sports facilities were commandeered), with massive audiences, many students but evangelicals shipped in from all over the state. Morris and Gish (the latter particularly) were skilled debaters, quite equaling President Trump in their cavalier attitude towards facts, and realizing that sometimes (often) a good joke is worth hours of laborious technical explanation. Evolutionists, unused to this kind of format, would still be in their preliminary remarks, when they would be cut off, time expired. It did not help their cause that often they would get very irate when this happened.

Finally, things came to a climax in 1981. A young Creationist, who was also a lawyer, had written up a proposed bill, insisting on balanced treatment between evolution and Creation Science, and in the legislature of the State of Arkansas he found takers. It was proposed and passed at record speed, taking only one Friday afternoon, when most had left or were eager to leave. Bill Clinton had been governor of the state from 1978 to 1980, when, for the first and last time not minding his fences, he was kicked out of office. He returned in 1982, and continued as governor until 1992, when he defeated incumbent G. H. W. Bush and became President of the USA. In the interregnum, from 1980 to 1982, was a man (Frank D. White) who was as surprised to find himself governor as he was unfit for the post. Unreflectively, he signed the bill, and on March 19, 1981, Arkansas Act 590 became effective. (Ruse 1988 gives the Act and the judge's opinion).

2. *The trial and its underpinnings*

As with *Epperson versus Arkansas*, the ACLU swung into action, preparing to bring suit against the law on account of its unconstitutionality. It lined up an impressive number of Arkansas religious leaders as plaintiffs, the lead being the Reverend William McLean, a United Methodist minister, whose name therefore became part of the subsequent trial and judgment—*McLean versus Arkansas*. (Actually, technically, *McLean* et al.*, versus The Arkansas Board of Education*). As is its wont, the ACLU looked for help from a prominent law firm, and the New York firm of Skadden, Arps, Slate, Meager and Flom came on board, pro bono, giving the free support of a rather junior female partner and a number of (very sharp) even-younger associates. (No one in the New York world of law is a disinterested altruist. This was very good publicity for the firm.) Everyone headed for trial, which took place in the first week of December 1981.

This is where I came in, in the early fall of 1981. Why me? I was not an American (not then, in 2000 I moved to a job in the States and ten years later became a citizen) and was not particularly distinguished. I was a (full) professor, fairly young (41), in the philosophy department of a university in Guelph, Ontario, Canada. It was not a major established university, having been founded only fifteen years previously, adding arts and sciences to the already existing Ontario Agricultural College, the Ontario Veterinary College, and McDonald Institute, the domestic science college. (In the early years, in my classes I often had students whose grandparents had met at Guelph).

I had however the background, the talents, and the eagerness that the ACLU was looking for in its search for expert witnesses to testify at the trial. In building its slate of witnesses, the ACLU had turned to one of the partners in Skadden, Arps, who was a trustee at Princeton. He phoned the president who put him onto several pertinent senor faculty. Then one of the young associates was given the job of talking to them. On the one hand, he wanted to get their take on the situation. What are the pertinent factors for instance, what kind of people should testify, what should be the overall strategy, what points needed emphasizing and what points needed avoiding? If you wonder how an untutored young lawyer could handle this job, let me say I was incredibly impressed with their intelligence (and diligence) and (perhaps as part of their training) they could soak up and conceptualize an area of knowledge and expertise. I still think that within a week or so, they could get to know 85% of the pertinent material in a field—and, undoubtedly, within a week of the end of the trial, forget it all!

On the other hand, the associate wanted the names of potential experts and witnesses, so at the end of every conversation he would ask for the names of two or three people and then set about phoning them, and in turn getting names from them too. Within a very short while, the basic, required strategy became apparent. You needed to go on the offensive with science witnesses, obviously, but also it was going to be very important to have theologians and other religiously knowledgeable (historians, sociologists) people to complement the science. And clearly you were going to need educators, those knowledgeable about the field and the issues, but also just plain classroom teachers who would explain how things happen in the classroom and how the balanced treatment law was simply wrong. Not the sort of thing to influence or shape the teaching of young people.

Soon, expectedly, certain names kept coming up again and again, and the eventual witnesses practically chose themselves. There was Langdon Gilkey, professor at the Chicago Divinity School, and the leading Protestant theologian in the country. There was George Marsden, evangelical historian of religion, then at Wheaton College, later at Notre Dame. There was—this was a foregone conclusion—Stephen Jay Gould at Harvard, evolutionist, and one of the best-known scientists in America because of the monthly column he wrote for the science magazine, *Natural History*. There was Francisco J. Ayala, Spanish-born, former priest, now one of America's most distinguished evolutionary geneticists. And there were more, including Arkansas school-teachers. (Missing was Carl Sagan, the most famous scientist of the day. He had been a little hoity toity when first approached. Later, as the approaching trial started to gather publicity, he offered his services. But it was too late).

But why me in this August group? Obviously, my name had come up, so I was not entirely unknown, and there was reason for this. I was one of a number in the 1960s (prominent member, David Hull from Chicago) who had kick-started the modern sub-field of the philosophy of biology, leading to my writing an introduction to the area, *The Philosophy of Biology* (1973). Also, like many in the 1960s, I had been much intrigued by *The Structure of Scientific Revolutions* (1962) by Thomas Kuhn. It was not so much that I was taken by his thesis of change—more on this in a moment—but that I was excited by his demand that philosophers of science take seriously the history of science. So much so, that I took my first sabbatical (1972–1973) in Cambridge, England, working in the University Library, immersed in the Darwin Archives. This led to my writing *The Darwinian Revolution: Science Red in Tooth and Claw* (1979). I joke that this is the book I wish I could have read ten years previously, when I was just getting into the field. But it is not really a joke. In a way, it is the complement in the history of science—the history of evolutionary biology particularly—to my *The Philosophy of Biology* in the philosophy of science—the philosophy of evolutionary biology particularly. It is a full overview of the revolution, making use of twenty years of archival research by Darwin scholars, including myself. I called it an "overview" in my preface. I expected all of the reviewers to say "No, no, Mike, it is much more than that. It is an original piece of scholarship." I didn't then realize that reviewers only read the first couple of pages of the book they are reviewing. Overview I said, overview they said, overview it is.

The point is that this pre-adapted me to take on the Creationists. It was not so much that I had done much work on the Creationist literature—although I had started work on this and by the time of the trial had a manuscript of what came out next year as *Darwinism Defended: A Guide to the Evolution Controversies.* I should say that the manuscript was circulated to both sides and became a major source for the state in my cross-examination. What I had done is much work on the kinds of arguments that the Creationists used. Many of these arguments were not that new and were around (and answered) at the time of Darwin. I knew the ropes. In fact, I had already a year or two earlier debated Morris and Gish in the basketball arena of Northwestern University in Chicago. By then I had over ten years of undergraduate teaching—a lot of it!—under my belt, so I was confident on my feet and I too knew that a good joke is worth ten arguments. I cannot say that I and my biologist partner won the debate. There must have been three thousand in the audience, at least ten of whom were evolutionists. But, within seconds of getting on the podium, I realized that this was my kind of event, I had great fun, and I saw my partner make all the mistakes—not getting to the main verb before he was cut off—and knew how to avoid them.

I had background preparation, I had the kind of personality that made me a natural for this sort of thing, and I was eager to do it. Not just the publicity—although most of my relatives and friends would say it was all about the publicity—but because I really do have moral concerns. As someone raised a Quaker, for all that my beliefs were long gone, I worried a lot about whether what I was doing was worthwhile, serving my fellow humans. You might say that of course being a teacher means you are serving your fellow humans, and increasingly over my life I came to see that. I really enjoy the scholarship, but I do take teaching seriously and have done my share and more. I have been at it for fifty-five years, and go on not just because, having married one of my students who was over twenty years younger than I and had three more kids, I need the money! But it wasn't like being a doctor, for instance, where so obviously you are serving others. I rush to say I don't want to be a doctor, although the first week or two of being a gynaecologist might be fun. Then the Arkansas trial came along and I saw a real chance of getting up and fighting what I believe are wrong and socially dangerous beliefs. It is not me and my pals who are against abortion, against homosexuals, and don't think women should be ordained. I should say that combined with this was the fact that my fellow philosophers wanted nothing to do with any of this. They thought it vulgar and misplaced to get into the witness box. Philosophers are not like other men. (Some had more legitimate concerns. David Hull was gay at a time when homosexuality was still much in the closet. He didn't want that coming out and being used, publicly, to discredit him, in a court trial).

l got roped in, and, in the fall of 1981, went off down to New York City to be deposed before the trial. It was then that I discovered that the lawyers for Skadden Arps were by no means convinced that I should be a witness. It was not so much me as generally a prejudice against philosophers. We tend to go on and on about arcane topics, that no one can or wants to understand, and on top of that we are so very arrogant. Convinced to a person that we are the brightest people on campus, we don't take instruction very well. You soon learn that lawyers are less concerned about the truth than about winning and this can lead to some very tense times. On top of this, of course, why a philosopher? Obviously, you need scientists, and theologians need hardly more justification. Educators are a must, and if you want to round things out with a historian or like person, why not? But why a philosopher?

As it happens, this had nothing to do with my merits. The Creationists rather forced it on the plaintiffs. It is a big mistake to think that Creationists are necessarily stupid – before he changed track, Gish had published in the *Proceedings of the National Academy of Science* (Gish et al. 1960)—and they certainly do their homework. They knew full well that the biggest thing to hit the philosophy of science in the past half-century had been Thomas Kuhn's *The Structure of Scientific Revolutions.* (Before Popperians who are reading this essay throw it down in uncontrollable rage, as belittling the status of their hero, note what I am saying and more importantly what I am not saying. I am not saying *The Structure of Scientific Revolutions* was the most important book or the most profound book or the longest-lasting book. I am talking about immediate attention and controversy, and Kuhn's book wins hands down).

The Creationists had studied *Structure* with great care and they knew full well the central concept and its supposed implications. Paradigms! Those conceptual frameworks within which scientific thinking is embedded. And what is the biggest mark of a paradigm, that which makes it so different and so controversial? That commitment to paradigms and changing from one to another is not simply a matter of reason and evidence. Paradigms require a kind of commitment to be found in religion or politics. People do change from being, say, a Catholic, to being a Protestant. Luther did! And people go the other way. John Henry Newman for example. But the change from one to the other is not simply a matter of sitting down and saying "I prefer consubstantiation to transubstantiation" or "I'm into justification by faith rather than good works." These may be important factors but in the end they are not decisive. Change needs almost a Kierkegaardian leap of faith. Creationists seized on this and argued that Darwinian evolution and Creation Science are different paradigms, with the supposed implication that one is as good as the other, and you cannot impose choice from without. At this point, you go beyond rationality and so that is it. There is no justification in education for preferring evolution over Creationism. Balanced treatment is not only the fairest moral way forward, it is sanctioned by strong (and fashionable) philosophical argument.

How were our lawyers—as I will now feel free to call them—to counter this? They too were bright and had done their homework. They knew full well that when Kuhn came onto the scene, and started to pick up steam in the mid-sixties, the person and the group most immediately and strongly in opposition were Karl Popper and his merry men. Above all, as spelt out in his *The Logic of Scientific Discovery* (first published in English in 1959), Popper stood for rationality and, above all, he found it in science. What separates science from all else is the demarcation criterion of falsifiability. Even the best science is constantly putting itself to the test of the empirical evidence and, if it cannot handle this, it falls. No matter how prestigious. The way that Newtonian mechanics—the best and most fruitful science ever—had had to give way before Einstein and the other physicists of the twentieth century. Kuhn is wrong. Call them paradigms or whatever, if they are part of science, they must be falsifiable. Science is not like religion. And if you doubt that, go and look at the book edited by Imré Lakatos and Alan Musgrave, *Criticism and the Growth of Knowledge* (1970), the report on a conference earlier in the decade, where the philosophies of Popper and Kuhn were spelt out and the two sides went at each other, trying to show the flaws in the position of their opponents.

The urgent need of a philosopher became obvious and the argument that the philosopher must make was no less obvious. The Kuhnian strategy must be countered and Karl Popper showed the way! I became part of the team that descended on Little Rock Arkansas, in early December 1981.

3. *The trial*

This was what was at stake:

On the side of Creation Science the claim was:

Sudden creation of the universe, energy and life from nothing;

The insufficiency of mutation and natural selection in bringing about development of all living kinds from a single organism;

Changes only with fixed limits of originally created kinds of plants and animals;

Separate ancestry for man and apes;

Explanation of the Earth's geology by catastrophism, including the occurrence of worldwide flood;

A relatively recent inception of the Earth and living beings.

On the side of evolutionary science, the claim was:

Emergence by naturalistic processes of the universe from disordered matter and emergence of life from nonlife;

The sufficiency of mutation and natural selection in bringing about development of present living kinds from simple earlier kinds;

Emergency [sic] by mutation and natural selection of present living kinds from simple earlier kinds;

Emergence of man from a common ancestor with apes;

Explanation of the Earth's geology and the evolutionary sequence by uniformitarianism; and

An inception several billion years ago of the Earth and somewhat later of life.

(Act 590 in Ruse 1988)

Naked mud wrestling! Moses versus Charles Darwin. Less exuberantly, are we faced with two co-equal paradigms, or are we faced with a religious claim and a scientific claim?

The actual trial, in a federal court, before judge William R, Overton, appointed to the post a couple of years earlier by President Jimmy Carter—Overton died in his forties later in the decade, I have often wondered if he might have been appointed to the Supreme Court by President Bill Clinton—took about a week and a half, as is normal, with the plaintiffs going first. The first day was given over to the people with religious qualifications—highly impressive was Langdon Gilkey, who took pleasure in pointing out all the Christian heresies being committed by the Creationists. (Interviewed later by Edward J. Larson, Overton said that it was Gilkey's testimony that started the downward slide of the Creationism case. I can well believe that.) The second day was given to the scientists. One point of note was that the state could not wait to get Stephen Jay Gould off the witness stand. He was somewhat chagrined, but one can understand they did not want to tangle with him. The third and final day was for the educators. The school-teachers were very moving. These people cared about their kids and their welfare.

I was slotted in on the second day in the morning. I was on the stand all morning and called back for a few minutes in the afternoon. I therefore had at least twice as much time as anyone else—my direct testimony was only half an hour, so cross examination was the best part of three hours. It was clear right from the beginning that the state's prosecutors thought, that if there was going to be a weak point, it was going to be the testimony of a philosopher. We are so out of touch with reality so much of the time! As it happened, it all went smoothly and, if I got too carried away with the sound of my own voice, our side would jump up, intervening, and letting me know that enough from me was twice as much as was needed! I am still proud of my one big joke. The assistant district attorney was harping on about my religious beliefs, trying to show that I am an infidel and so what would you to expect me to say about evolution? Eventually, frustrated, I blurted out: "Can't you see Mr. Williams, I am not an expert witness about my own religious beliefs." Everyone laughed and, when Williams tried to continue that line, the judge intervened and told him to move on. "Can't you see, he's not going to give you what you want."

Expectedly, both plaintiffs and defense made much of Popper. I opened this essay with what I said to our side early in the morning, and under cross-examination we came back to it again and again. But really, that was easy. We had a party line and stuck to it. Evolutionary theory can be falsified and Creation Science cannot be. This is a good example of the sort of thing that went on under cross-examination:

Q: You've talked about how the creation scientists quote evolutionists out of context, using one sentence. Yet, if an evolutionist should quote a creation scientist out of context, would that be any less dishonest, in your opinion?

A: I think that I would have to say that it would be no less dishonest if one sort of played fast and loose with that point there.

Q: And when you quote from some of the books you mentioned earlier, specifically, Doctor Gish's book, you didn't point out to the Court, did you, that Gish goes on to talk about how neither, under the pure definition as articulated by Karl Popper, neither evolution nor creation science can qualify as a scientific theory?

A: I thought it was—

Q: Did you point that out? If you did, I didn't hear it.

A: Well, if you didn't hear it, then I expect I probably didn't. But I, you know— Let me add very strongly that I want to dispute the implication that I'm being dishonest at this point.

My understanding was it wasn't evolution on trial here; that it was, if you like, creation. That's the first point. And secondly, as you know, I personally don't necessarily accept everything that Popper wants to say. So I've don't think that I've quoted Gish out of context at all. I was asked to give an example of a passage in scientific creationist writings where the scientific creationists quite explicitly appeal to processes outside the natural course of law.

Now, I'd be happy to reread it, but I think that's what I did, and I think I did it fairly.[3]

(I must say that, rereading this stuff forty years later, I am quite impressed with my poise. I am not sure that today I am quite that self-confident about anything! But then, for nearly forty years, I have been married to a wife much younger than I).

I left Arkansas after our side had finished testifying. In a way, I felt a bit sorry for the Creationists. We had such a stellar cast (I am not talking about me). They really had to scrape the barrel. No Langdon Gilkeys or Steve Goulds for them. Judge Overton handed down his ruling in early January in 1982, and it was unambiguous. Evolution is science. Creationism is religion. Teaching the latter violates the First Amendment separation of Church and State. "The Act was passed with the specific purpose by the General Assembly of advancing religion." No balanced treatment for the kids of Arkansas. (See Ruse 1988).

The points that the judge made were all fairly obvious and expected. No one in the real world ever accused him of misreading things or getting into dubious convoluted arguments. Again, at the risk of seeming unduly immodest, my testimony was at the heart of his ruling.

It is guided by natural law;
It has to be explanatory by reference to natural law;
It is testable against the empirical world;
Its conclusions are tentative, i.e. are not necessarily the final word; and
It is falsifiable.

> In making these points, explicitly the judge referenced me. "Ruse and other science witnesses."

I should say that, as things went, I don't think there was any big surprise that my testimony turned out to be so central. I have said that, for a long time, our attorneys were not at all convinced of the wisdom of using a philosopher. Indeed, even on the Sunday, the day before the trial, by which time I had flown to Arkansas, there was discussion about whether to use me. It really wasn't me personally that was at issue—although after the final rehearsal the night before my testimony, my attorney said "Finally Mike, I think you are doing a better job than I could do." I didn't take that as a criticism. The big question was the same all along. Could one risk putting a philosopher on the stand? I was neither fish nor fowl, and, as the state attorneys showed in spades the next morning, if they were going to be able to tear holes in the plaintiff's' case, the wild and wooly thinking of a philosopher was just the place to start. They weren't going to take on someone like Steve Gould. However, once the decision had been made to use a philosopher, then it immediately became clear to everyone that this was going to be the make or break testimony. Could one show that Creation Science is religion? Could one show that evolution is science? These are philosophical questions and if you get them right, you can win. We did get them right, and we did win.

Before I get to the aftermath of my testimony, let me give a bit more history. The state did not appeal the ruling so, technically, it only applied to a certain part of Arkansas. However, a year or two later, a similar case came up. *Edwards v. Aguillard*, in the State of Louisiana.[4] I was deposed for that case too (so I had obviously not blotted my copy-book in Arkansas!) and so this time was Carl Sagan. However, it never went to trial and was rapidly moved up the greasy pole. It went all the way to the Supreme Court, where once and for all teaching Creationism was ruled a breach of the First Amendment separation of Church and State. Although of course these things never are once and for all. By 1990, Creationism morphed into the more user-friendly Intelligent Design Theory (IDT). In the first decade of this century, a school board in Pennsylvania no less—not one of the expected evangelical states of the South—opened the possibility of teaching IDT. Quickly the ACLU got involved and it came to trial—*Kitzmiller versus Dover Area School District* (2005).[5] Again, the biblical side lost, after a ruling by a conservative judge (a 2002 appointee of George W Bush). For some years, I had given up writing about Creationism as such—it is politically important but intellectually rather boring—and so I neither expected to be nor was I asked to be a witness. I should say that some of my writings by then—of which more in a moment—surely convinced the ACLU lawyers that they should have nothing to do with me!

4. *The place of Popper*

Let me now concentrate on three follow-up matters. First, my testimony and the use of Popper. If you look at the ruling you see that Popper's criterion of demarcation—falsifiability—was not just a crucial part of my testimony, but a crucial part of the judge's ruling. So as far as winning was concerned, it was the right strategy. But was it true? I must say that I did not then and do not now think of myself as a Popperian, in the sense of thinking his work is so central to the philosophy of science that he was the most important philosopher of science of the twentieth century. Indeed, like most people I found intensely irritating the group of sycophants with which he surrounded himself. One of the most annoying experiences of my academic life ever was trying to give a paper with the broadcaster and writer Bryan McGee in the audience. Every time I tried to make a comment, he would spring to his feet and tell the audience that Popper had made a similar point in a rather better manner, or that my understanding of something involved an egregious misreading of Popper's philosophy. I was tempted to give the podium over to McGee and I fully expect he would have taken it.

That said, there is (or was) in US philosophy of science circles intense hostility to Popper and his ideas. I never shared this feeling nor do I now. I met Popper only two or three times, but when we did meet, our encounters were very cordial. In fact, I wrote a paper in the late seventies critical of his claims about evolutionary theory (Ruse 1977). When we met, he remembered it (and brought it up), said he thought I had a point (John Maynard Smith had really put him right on these issues), and we had what I thought was a very fruitful conversation. My long-time colleague, Tom Settle, once told me that Popper had difficult relations with his children (his students and the like) but got on really well with his grandchildren—of which Tom, as a student of Joe Agassi, was one. Most importantly, I had—and still have—huge admiration for Popper as a voice of rationality in the 1930s and 1940s, at a time when the world was in dire need of voices of rationality. That is by far my overwhelming emotion when I think of Karl Popper.

As far as the philosophy was concerned, it wasn't so much that I was in favor of falsifiability or against falsifiability. It was rather that it was never really a topic of mine. In the philosophy of science, I was working on theories and their construction—people like Hempel and Nagel were more central to me. Quite apart from the fact that I am not a physicist, so Popper's work was not really my flavor. Then, when I worked on the history of science, the philosopher I had in my targets was Kuhn, as I tried to show that the Darwinian Revolution could not have been as Kuhn hypothesized. There was no abrupt switch from one position to another—incommensurable paradigms—but a general gradual change, with Darwin's thinking incorporating much that he had learnt from the non—or anti—evolutionists. To this day, I say that Darwin was a rebel not a revolutionary.

The one exception to my lack of real interest was that already-mentioned paper on Popper on evolutionary biology. He had said that Darwinian theory is not real science but a metaphysical research programme that could not be falsified—apart from anything else, he claimed that natural selection is a tautology so obviously is not empirical (Popper 1974). The first paper I ever had accepted—a presentation at the first meeting of the PSA in 1968 in Pittsburgh—was on that topic. I guess I was interested in falsifiability in a minor way, right through practically until the Arkansas trial. I thought then as I think now that falsifiability is important and it is a mark of genuine science, although I was not then (nor am I now) convinced that that is all there is to be said on the topic of demarcation. Overton got me right. Falsifiability is important but there are other factors too. In Arkansas I was not selling my birthright for a mess of pottage, or, more prosaically, the chance to get involved in an exciting and very public event.

5. *A pariah among the respectable*

Second, for my testimony in Arkansas, I got it in the neck from my fellow philosophers. I was a bit surprised. I thought that, even if people didn't quite agree with me, there would be respect for what I had done. No way. The first intimation of how things were going to go was at the Eastern APA, just after Christmas 1981—after the trial but before the ruling. I knew Ernan McMullin—philosopher of science, Galileo expert, Catholic priest, professor at Notre Dame, Irishman—quite well and thought of him as a friend, for all that he was a generation older than I (and I am English-born). It was at the smoker—APA meetings used to have those sorts of things in those days—and I was a bit cocky about what I had just done in Arkansas—me and Steve Gould sort of thing. I was gobsmacked. Ernan went bright red and had trouble talking to me. I had demarcated science from non-science? And I had used Popper as my foundation? It was not a pleasant encounter.

I should say that a year or two later, Ernan McMullin and I were back on good terms and, after he died in 2011, in the science and religion journal *Zygon*, I wrote an appreciation of him and of my great philosophical debt (Ruse 2012). In a PSA Presidential Address, Ernan had taken up the question of epistemic values (prediction, confirmation, falsifiability) versus non-epistemic values (racism, homophobia, sexism) in science, arguing that over time the former expel the latter McMullin 1983). I don't think he was quite right, but his thinking spurred me to write what is perhaps my most important book, certainly my longest, *Monad to Man: The Concept of Progress in Evolutionary Biology* (1996). I argue that scientists kick out non-epistemic values not because they no longer believe in them but because their presence goes against the standards of good professional science, and above all scientists want to be considered good professionals. They are real scientists, not phrenologists or whatever. I am now fairly sure that what made Ernan so mad at that smoker was not at all the appeal to Popper, but that he thought I was simply attacking religion. Ernan trod a careful path between being a very secular philosopher of science—no Thomist he!—and a Catholic Priest, moreover a rather conservative Catholic Priest. Although he was for many years at Notre Dame, he was not a member of any order, and always under the suzerainty of his bishop back in Ireland (Eire). Some years later, Ernan gained as a colleague the Calvinist evolution-hater Alvin Plantinga and I thought Ernan became much more appreciative of my position. (By then I was quarreling with the New Atheists, so, although a non-believer, I was acknowledged far and wide as no rabid opponent of God and religion).

As happens with philosophers, things soon got into print. Another good friend (!) Larry Laudan went after me with hammer and tongs.

In the wake of the decision in the Arkansas Creationism trial (McLean v. Arkansas), the friends of science are apt to be relishing the outcome. The creationists quite clearly made a botch of their case and there can be little doubt that the Arkansas decision may, at least for a time, blunt legislative pressure to enact similar laws in other states. Once the dust has settled, however, the trial in general and Judge William

R. Overton's ruling in particular may come back to haunt us; for, although the verdict itself is probably to be commended, it was reached for all the wrong reasons and by a chain of argument which is hopelessly suspect. Indeed, the ruling rests on a host of misrepresentations of what science is and how it works. (Laudan 1982, 16; reprinted in Ruse 1988).

And that is just a warm-up. Basically, Laudan criticized me for offering criteria of demarcation, including falsifiability. And he thought that I was aiming at the wrong end. The question is not whether Creation Science is science but whether it is good science. It is bad science and so should not be taught in the classroom. Demarcation issues are side stepped.

The core issue is not whether Creationism satisfies some undemanding and highly controversial definitions of what is scientific; the real question is whether the existing evidence provides stronger arguments for evolutionary theory than for Creationism. Once that question is settled, we will know what belongs in the classroom and what does not. (ibid. 18).

Expectedly, falsifiability got roughed up.

Judge Overton was explicitly venturing into philosophical terrain. His obiter dicta are about as remote from well-founded opinion in the philosophy of science as Creationism is from respectable geology. It simply will not do for the defenders of science to invoke philosophy of science when it suits them (e.g., their much-loved principle of falsifiability comes directly from the philosopher Karl Popper) and to dismiss it as "arcane" and "remote" when it does not. However noble the motivation, bad philosophy makes for bad law. (ibid. 19).

My reply—and once again I am rather impressed at the confidence and robustness of what I thought and wrote—was, first, that the US Constitution does not forbid the teaching of bad science. It forbids the teaching of religion. It is no good trying to do an end run around demarcation criteria. Second, it is just silly to say that there can be no such criteria. Take a statement like "The Earth is flat." (I am using examples from now to make the point.) You cannot just work from marks on paper. Interpretation counts. Obviously, if you are prepared to accept empirical evidence, it is falsifiable. Go to the sea-shore, look at the horizon, and ask why ships coming towards land first show their masts and only gradually is all else revealed. But if you are not prepared to accept such evidence—you have religious reasons for holding always that the earth is flat—then your position is unfalsifiable. If you keep invoking things like optical illusions, then you are into religion not science. It is true that in my response to Laudan, I do rather lace into Popper. "Simple criteria that supposedly give a clear answer to every case-for example, Karl Popper's single stipulation of falsifiability will not do." (Ruse 1982, 21; reprinted in Ruse 1988) (To be fair, I am not sure that Popper ever thought this either.) But then I make it clear that I am not throwing Popper overboard. Anything but.

Finally, what about Laudan's claim that some parts of creation-science (e.g., claims about the Flood) are falsifiable and that other parts (e.g., about the originally created "kinds") are revisable? Such parts are not falsifiable or revisable in a way indicative of genuine science. Creation-science is not like physics, which exists as part of humanity's common cultural heritage and domain. It exists solely in the imaginations and writing of a relatively small group of people. Their publications (and stated intentions) show that, for example, there is no way they will relinquish belief in the Flood, whatever the evidence. In this sense, their doctrines are truly unfalsifiable. (ibid. 22)

Unlike Laudan, I had read the Creationist literature and could quote it.

… it is… quite impossible to determine anything about Creation through a study of present processes, because present processes are not created in character. If man wishes to know anything about Creation (the time of Creation, the duration of Creation, the order of Creation, the methods of Creation, or anything else) his sole source of true information is that of divine revelation. God was there when it happened. We were not there… therefore, we are completely limited to what God has seen fit to tell us, and this information is in His written Word. This is our textbook on the science of Creation! (ibid. 21).

This is not science. And if further proof is needed, look at the testament of faith that one had to sign in order to become a member of the leading organization, the Creation Research Society.

(1) The Bible is the written Word of God, and because we believe it to be inspired throughout, all of its assertions are historically and scientifically true in all of the original autographs. To the student of nature, this means that the account of origins in Genesis is a factual presentation of simple historical truths. (2) All basic types of living things, including man, were made by direct creative acts of God during Creation Week as described in Genesis. Whatever biological changes have occurred since Creation have accomplished only changes within the original created kinds. (3) The great Flood described in Genesis, commonly referred to as the Noachian Deluge, was an historical event, worldwide in its extent and effect. (4) Finally, we are an organization of Christian men of science, who accept Jesus Christ as our Lord and Savior. The account of the special creation of Adam and Eve as one man and one woman, and their subsequent fall into sin, is the basis for our belief in the necessity of a Savior for all mankind. Therefore, salvation can come only thru accepting Jesus Christ as our Savior (ibid. 22).

Enough said. Except a reflection of my thoughts then and my thoughts now. I am a professional philosopher. I love the attacks and counter-attacks that are part of our trade. And, I am certainly not averse to publicity. A few years after the Arkansas trial, I put together a collection—*But is it Science? The Philosophical Question in the Evolution/Creationism Controversy,* that includes material of historical significance (mainly articles by me), material of contemporary relevance (mainly articles by me), and follow up material (articles by me but also of my critics like Larry Laudan). "Therefore if thine enemy hunger, feed him; if he thirst, give him drink: for in so doing thou shalt heap coals of fire on his head" (Proverbs 25, 21–22).

My leading emotion however, then and now, was/is one of contempt. Creation Science is dangerous. It should not be taught in classrooms. We see only too well the pernicious effects of pseudo-science and like phenomena, including extreme evangelical religion. Anti-vaccination, anti-global warming, anti-GMOs—at a time when diseases run rampant, cities are lost under the sea, half of the world's children go to bed hungry. Laudan and his fellows had no thought for this. And before you protest that they were after truth not comfort, then why didn't they look more carefully at the philosophical issues at stake? Why didn't they spend even one afternoon looking at the Creationist literature? Even in the pre-internet era it was not hard to find. Whatever issues I had with Popper and his coterie of groupies, again I go back to the stand he took for rationality when it was so needed. That to me is a real Mensch.

6. *Darwinism as religion*

Third, let me conclude this essay—more a memoir!—by taking up the effect, by the Arkansas trial, on my subsequent professional career as a scholar. I could not other than be struck, at a kind of meta-level, at what was going on here. Why the hostility to evolution, especially to Darwinian theory and its mechanism of natural selection? Simplistically, because it goes against the bible. Yes, but no one hates the Copernican theory even though, supposedly, the sun stopped for Joshua. In any case, it is not a generic hatred. Creationists admit these days that the Ark would not have been big enough to carry all the species of animal extant today. Their ploy is that the Ark carried "kinds," and after the Flood these evolved into the different species we have today. And how did the evolution occur? Natural selection! The Creationist Museum in Northern Kentucky has a better display and discussion of natural selection than the Field Museum, in Chicago, 300 miles to the north.

I got the key insight giving the answer to my question from, of all people, the Creationist Duane T. Gish. I should explain I always had very good relations with the Creationists, Gish in particular. I guess we recognized fellow performers. I have in my possession a copy of *Evolution: The Fossils say No!* (over 150,000 copies sold), inscribed by Gish to Michael Ruse, his good friend, with warm best wishes. Hope springs eternal in the breast of this bibliophile that Gish will be proven right, and Darwin wrong, and I shall be the owner of a rare, much-coveted first edition. I also got on well with the State of Arkansas Attorney General. He was very smooth and later in the decade ended in jail for fraud. I am not surprised. And completing the list of my odd friendships, I was a good pal of the devisor of Intelligent Design Theory, Berkeley law professor Phillip Johnson. He was born on June 18, 1940, and I on June 21, 1940. I always joked that it showed that God had a sense of humor, to invent Phil and then me to give him ulcers. I contributed to his Festschrift. The same year I contributed to the Festschrift of leading New Atheist, Richard Dawkins. I don't really know about God's sense of humor. Mine is pretty active.

After the trial, I got to know Gish well as we appeared often together on TV talk shows, and a constant theme of his complaint about Darwinism was that it was really just as much a religion as Creationism. (We were off-stage, so he was quite happy making those judgments of Creationism.) For a long time, I resisted his suggestion, but then came to realize that he had a point. It is not so much that Darwinian theory is religion. That is perfectly good science in its own right. It is rather that people take Darwinism and use it as the basis for a form of secular humanism. Just think, we have Darwin Day, celebrating Darwin's birth. So, also, we have Jesus Day. We call it Christmas. We don't have Copernicus Day or Newton Day or, for that matter, Dawkins Day. (I suspect he would be embarrassed, but not that much).

This insight set me on a thirty-year journey, trying to show exactly how Darwin's theory is turned into a religion. I wrote a book showing that both Creationism and Darwinism (construed in this sense) are into eschatology, world systems about meaning and end times (Ruse 2005). Creationists are Providentialists, thinking we can do nothing without the saving grace of the blood of the lamb, and so we must prepare for the end trying to obey His commands. Darwinists are progressionists, thinking we must improve things through our own efforts, if we are to bring Jerusalem down here on Earth. Both sides are into heaven—a secular version at least for the Darwinists—but they have different prescriptions on how to achieve it. In the lingo of theology, Creationists are pre-millennialists, thinking Jesus will come before things are put right. Darwinists are post-millennialists, thinking Jesus (in a metaphorical sense) will come later when we have put things right. More recently, I wrote a book on Darwinism and literature, showing how fiction and poetry show that folk worked through such Christian themes as origins, God, the status of humans, sin, sex, salvation from a Darwinian perspective (Ruse 2017). I followed this with a book on war, showing how Christians and Darwinians took different stances on all the moral issues that such conflict entails (Ruse 2018).

What was fascinating was how, topic after topic, I found parallel treatments. Both Creationists and Darwinists obsessed about the special place of humans, for example, determined to find that everything revolves around us—made in the image of our Providential God as opposed to the climax of a progressive process of evolution. Showing that there is a lot more than just science going on here, the scientific theory of Darwinian evolution explicitly eschews such progress. Humans are different, but the science does not say we are better. In fact, the opposite. In the immortal words of the paleontologist Jack Sepkoski: "I see intelligence as just one of a variety of adaptations among tetrapods for survival. Running fast in a herd while being as dumb as shit, I think, is a very good adaptation for survival" (Ruse 1996, 486).

I will not labor the point. If you are interested, I have written extensively—very extensively—on the topic. What I will note is that my claims in this sphere are the strong reason why I am the last person that the ACLU wants up on the witness stand. Imagine when the defense attorneys get going on my claims about Darwinism being a religion. The fact that I have always insisted that there is a genuinely scientific Darwinian theory of evolution will be regarded as an irrelevant joke. I should say I am not being paranoid. There is good evidence for my suspicion. The Creationist Paul Nelson (another good friend!) took note of a AAAS meeting in 1993, where I gave a talk. Nelson remarks:

Michael Ruse, a philosopher and biology historian at the University of Guelph in Ontario, was probably the best-known speaker featured at the session, "The New Anti-evolutionism." As session organizer Eugenie Scott remarked before Ruse spoke, "He is almost a person who needs no introduction in this context." Yet a recent article describing the session in the *London Times* Higher Education Supplement omits Ruse entirely. Although the *Times* provides the identities and views of all the other speakers in some detail, they make no mention–even in passing–of Ruse nor his talk.

Why the glaring omission? Was Ruse's talk so commonplace or forgettable that it warranted no mention? Hardly: indeed, the opposite is the case. Ruse is often controversial, but he is rarely boring, and his talk entitled "Nonliteralist anti-evolution as in the case of Phillip Johnson" was true to form; it was (for this correspondent) easily the most memorable and surprising of the meeting. Thus I speculate that Ruse's conspicuous absence from the *Times* article may be due to a certain uneasiness about his main point, which, Ruse argued (and I agree) "is an important one."[6]

Looking at what I said, Nelson had a point. He records my talk covering an earlier encounter I had had with Phillip Johnson, I said.

What Johnson was arguing was that, at a certain level, the kind of position of a person like myself, an evolrutionist, is metaphysically based at some level, just as much as the kind of position of…some creationist, someone like Gish or somebody like that. And to a certain extent, I must confess, in the 10 years since I performed or I appeared in the creationism trial in Arkansas, I must say that I've been coming to this kind of position myself.

Nelson picks up the thread:

It is now important, Ruse continued, that evolutionists admit–to themselves, if not "in a court of law"–that "the science side has certain metaphysical assumptions" which ground its view of origins, and that future discussions must take account of these assumptions. We cannot ignore them.

One problem is that the picture of science received from the "logical positivists" or "people like Popper and Hempel and Nagel" accords poorly with much historical evidence concerning evolution's role. "It's certainly been the case," Ruse said, "that evolution has functioned, if not as a religion as such, certainly with elements akin to a secular religion." As examples, he cited "the most famous family in the history of evolution, namely, the Huxleys," and, more recently, biologist Edward O. Wilson. About Thomas Henry Huxley, Darwin's "bulldog," Ruse noted:

Certainly, if you read Thomas Henry Huxley, when he's in full flight, there's no question but that for Huxley at some very important level, evolution and science generally, but certainly evolution in particular, is functioning a bit as a kind of secular religion.

Julian Huxley, Thomas's grandson, also stood in this tradition.

For many evolutionists, Ruse continued, things are much the same today: "Evolution in a way functions as a kind of secular religion."

In his book *On Human Nature*, well-known Harvard systematist and sociobiologist E.O. Wilson "is quite categorical," he argued, "about wanting to see evolution as the new myth, and all sorts of language like this. That for him, at some level, it's functioning as a kind of metaphysical system."[7]

I guess with friends like me, you don't need any enemies! More seriously I stand by every word I said. I stand also by my lifelong commitment to Darwinian theory as hugely important science and a testament to the real reason why we are made in the image of God, whether or not He exists. This was true back at the time of the sociobiology debate, when I took a huge amount of flack for my conviction that Edward O. Wilson was right in seeing human nature as a product of Darwinian evolution (despite his yearnings for something more), to my recent arguments about the Darwinian evolution of morality and its consequent ontologically non-real status.

Envoi

Although I have not in any sense been working in a Popperian mode, you can see how the whole demarcation (of science from non-science) issue has permeated my intellectual being and drives forward the work I do. Science, religion, and the differences between them. This why I can conclude that, although I am not in any recognizable (or non-recognizable) sense a Popperian, I am very glad and proud to be in the same intellectual field as he.

Notes

1. http://www.antievolution.org/projects/mclean/new_site/pf_trans/mva_tt_p_ruse.html.
2. http://cdn.loc.gov/service/ll/usrep/usrep393/usrep393097/usrep393097.pdf.
3. http://www.antievolution.org/cs/mclean_ruse_test.
4. http://cdn.loc.gov/service/ll/usrep/usrep482/usrep482578/usrep482578.pdf.
5. https://law.justia.com/cases/federal/district-courts/FSupp2/400/707/2414073/. A second edition of *But is it Science?* was co-edited by Robert Pennock who was the philosophy expert witness in Dover. We include material on this trial as well as Arkansas. See Pennock and Ruse (2008).
6. http://www.arn.org/docs/orpages/or151/151meta.htm.
7. Ibid.

References

Biological Sciences Curriculum Study.: *Biological science: molecules to man.* Boston, Houghton Mifflin (1963)

Gish, D.: *Evolution: the fossils say no!* San Diego, Creation-Life (1973)

Gish, D.T., Tsugita, A., Young, J., Fraenkel-Conrat, H., Knight, Stanley, W.M.: The complete amino acid sequence of the protein of tobacco mosaic virus. Proc. Natl. Acad. Sci. **46**, 1463 (1960)

Kuhn, T.: The structure of scientific revolutions. Chicago, University of Chicago Press (1962)

Lakatos, I., Musgrave, A.: Criticism and the growth of knowledge. Cambridge, Cambridge University Press (1970)

Larson, E.J.: Summer for the gods: the scopes trial and america's continuing debate over science and religion. New York, Basic Books (1997)

Laudan, L.: Science at the bar-causes for concern. Sci., Technol., & Hum. Values **41**, 16–19 (1982)

McMullin, E.: Values in science. In: Asquith, P.D., Nickles, T. (eds.) PSA 1982, 3–28. East Lansing, MI: Philosophy of Science Association (1983)

Numbers, R.L.: The creationists: from scientific creationism to intelligent design. Cambridge, MA, Harvard University Press (2006)

Pennock, R., Ruse, M. (ed.): But is it science? the philosophical question in the creation/evolution controversy, 2nd ed. Buffalo, NY, Prometheus (2008)

Popper, K.R.: The logic of scientific discovery. London, Hutchinson (1959)

Popper, K.R.: Darwinism as a metaphysical research programme. In: Schilpp, P.A. (ed.) The philosophy of Karl Popper, 133–143. LaSalle, IL: Open Court

Ruse, M.: The philosophy of biology. London: Hutchinson (1973)

Ruse, M.: Karl Popper's philosophy of biology. Philos. Sci. **44**, 638–661 (1977)

Ruse, M.: The Darwinian revolution: science red in tooth and claw. Chicago, University of Chicago Press (1979)
Ruse, M.: Pro judice. Sci., Technol., & Hum. Values **7**, 19–23 (1982)
Ruse, M. (ed.): But is it science? the philosophical question in the creation/evolution controversy. Buffalo, NY, Prometheus (1988)
Ruse, M.: Monad to man: the concept of progress in evolutionary biology. Cambridge, MA, Harvard University Press (1996)
Ruse, M.: The evolution-creation struggle. Cambridge, MA, Harvard University Press (2005)
Ruse, M.: Science and values: my debt to Ernan McMullin. Zygon **47**, 666–685 (2012)
Ruse, M.: Darwinism as religion: what literature tells us about evolution. Oxford: Oxford University Press (2017)
Ruse, M.: The problem of war: Darwinism, christianity, and their battle to understand human conflict. Oxford, Oxford University Press (2018)
Whitcomb, J.C., Morris, H.M.: The genesis flood: the biblical record and its scientific implications. Philadelphia, Presbyterian and Reformed Publishing Company (1961)

Michael Ruse Lucyle T. Werkmeister Professor of Philosophy at the Florida State University, USA (retired) and Professor Emeritus, University of Guelph, Ontario, Canada is a historian and philosopher of science, with a strong interest in Charles Darwin and his theory of evolution by natural selection, both in history and today. He is the author of many books, including *Darwinism as Religion: What Literature Tells Us About Evolution* (OUP, 2017); *On Purpose* (Princeton UP, 2018); and The *Problem of War: Darwinism, Christianity, and their Battle to Understand Human Conflict* (OUP, 2018). E-mail: mruse@fsu.edu

Cognitive Science

Popper on the Mind-Brain Relation

Peter Århem

1 Introduction

This chapter examines the view of Popper on the mind-brain relation in the light of neuroscience. It is based on two interviews that Ingemar Lindahl and I made with him 1992 and 1994. In these interviews Popper presents an extension of his interactionist view, taking his point of departure from the observation that mind has many similarities with forces. The chapter is organized in three parts. The first discusses Popper's interactionism and competing views. His argument for an interactionist solution of the mind-brain question based on the theory of evolution is addressed, as is his view that biology is not reducible to physics. His interactionism is put into the context of the philosophical landscape of today, which is dominated by parallelist positions. His critical view on these positions are discussed, as well as his responses to the arguments against an interactionist position.

In the second part of the chapter his new view on mind is addressed. An interpretation suggesting that electromagnetic fields of the brain are an intermediate link between the conscious mind and the neuronal activity is presented. It is pointed out that the introduction of such an intermediate link that have properties in common both with conscious mind and with the spatio-temporal pattern of nerve impulses, may make it easier to conceive of an interaction between mind and brain. The chapter also addresses Popper's view that the relative autonomy of forces may be a relevant factor in the attempts to better understand the mind-brain issue. Further it discusses Popper's suggestion that the all-or-nothing principle of neurophysiology may be relevant for understanding how microscopic effects on the brain are amplified.

P. Århem (✉)
Department of Neuroscience and Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
e-mail: peter.arhem@ki.se

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_15

The final part of the chapter explores the idea that mind affects quantum mechanical probability fields. Two hypotheses are discussed. One is the microsite hypothesis of Beck and Eccles, assuming that the critical targets are cortical synapses. The other assumes that the targets are ion channels of cortical nerve cells. In both cases quantum tunneling is assumed to be involved. Quantitative estimations suggest that quantum mechanical principles may indeed play a role at a macroscopic level of cortical action. This is discussed with reference to Popper's view that a mechanism based on quantum principles is only one of many possible mechanisms to explain the mind-brain interaction.

2 Background

Popper had a long-standing interest in the mind-brain relation (Popper 1953, 1955, 1972, chapter 6, 1973, 1976, 1977, 1978, 1994; Popper and Eccles 1977). His most thoroughly discussed presentation is in "The Self and its Brain" from 1977, written together with the neurophysiologist John Eccles. In all his publications about this issue he argued for an interactionist solution of the problem, criticizing the dominant materialistic positions. Already from the beginning this was a controversial standpoint, and still is.

In my eyes his most convincing argument is based on the theory of evolution. It states that mind is a result of evolution, a result of biology, and not of physics. And biology is not reducible to physics. This is another controversial standpoint of Popper, epitomized in his dictum that biochemistry is not reducible to chemistry; a statement made in the debate with the chemist Max Perutz after his delivery of the Medawar Lecture in 1986 on active versus passive Darwinism (Niemann 2014). This of course entails that biophysics is not reducible to physics. In all areas of biology, aims and intentions are involved—and these are not used or needed in physics and chemistry.

> I do think mind is a result of biological evolution. It is something very special to biology.... and not to physics alone. Biology is something absolutely marvellous. Who, who saw the world before life arose in the world, who would have dreamt of all these marvellous … I think life is certainly the great turning point in the evolution of the world. We must admit that we know extremely little, and that materialism, I think, is just a word. And it really is refuted at every minute and every day of our lives. (Popper et al. 1994)

3 The Philosophical Landscape

Interactionism is a rather small, but not insignificant, stream in the contemporary philosophical landscape of today. There are other streams of thought that are more popular, at least in academia. The terrain can be mapped as follows:

1. a radical immaterialism, which denies that a material reality exists,

2. a radical materialism, which denies that consciousness exists,
3. a psychophysical interactionism, which assumes that conscious processes and neuronal processes interact and
4. a psychophysical parallelism, which assumes that conscious processes and neuronal processes are parallel, but do not but not interact. To this we can add
5. epiphenomenalism, which forms a hybrid between parallelism and interactionism, and which assumes that the brain unilaterally produces consciousness.

The two radical positions (1) and (2) are rather uninteresting from a neuroscientific perspective, as well as from Popper's perspective. The radical immaterialism simply assumes that material reality is merely a mental construction based on sense impressions. This was Berkeley's and Mach's solution of the consciousness problem. Radical materialism assumes that mental reality is merely behavior (Dennett 1991). The quality of consciousness does not exist. This position seems in a neuroscience perspective even less fertile than radical immaterialism.

The three other positions seem more fruitful. Psychophysical interactionism assumes that mental and neural processes in some sense interact. This was Descartes' classical solution and it is the common-sense solution in the light of the evolutionary theory, also adopted by Popper (Popper and Eccles 1977).

Psychophysical interactionism has a long history. It was not invented by René Descartes, as many mind-brain theorists would like us to believe (e.g. Ryle 1949). Interactionism can be found in prehistoric cultures (Solecki 1971) and it is a common view among present-day indigenous populations (see Bloch 2013; Descola 2013). Popper even asserted that all major thinkers before Descartes, and of course Descartes himself, were dualist interactionists in some form or another (Popper and Eccles 1977, 152). For Descartes the essence of matter was extension and for mind it was non-extension. The contact between the two was assumed to be found in the pineal gland, a rather reasonable idea since nerve activity was assumed to be hydrodynamic waves in fluids in hollow nerves and the fluid reservoirs in the brain are found in the system of ventricles in close contact with the pineal gland. Thus Descartes suggested that a system of valves in the pineal gland could be regulated by mind processes, under the precondition that the momentum of the fluid was preserved.

Psychophysical parallelism assumes that mental and neural processes perfectly follow each other, in a perfect 1 to 1 mapping. It was such an idea that was suggested by Leibniz and by Spinoza as a solution to the contradictions found in Descartes' interactionist theory. It is variants of this position, perhaps mainly the identity theory, which today dominates the philosophical landscape (Feigl 1967; Edelman 1992; Crick and Koch 1990; Searle 2004). It should be pointed out here, that Descartes view was based on rather detailed biological ideas about the anatomy of the brain, while Leibniz thinking was much more abstract. I think this is an observation of some interest here, related to discussion of instrumentalism by Popper in "Conjectures and Refutations" (1963). He there severely criticized instrumentalist attitudes, using as example the trial of Galilei and the conflict between the successful realistic ideas of Galilei and the unsuccessful instrumentalism of Roberto Bellarmino.

Both Spinoza and Leibniz advocated a specific form of parallelism with ancient roots, panpsychism; a theory assuming that everything materially also has a mental “inside” or internal aspect. This view has recently appeared in new shapes and seem to have gathered new adherents among neuroscientists (Smith 2008; Koch 2019). But still it seems as if the dominant parallelist positions today are different versions of the identity theory, assuming that conscious processes are in some sense (but not logically) identical, and thus in some sense (though not mathematically) completely parallel, with certain neural processes in certain parts of the brain.

The hybrid position of epiphenomenalism was originally presented by Thomas Huxley, one of Darwin’s friends and the first public defender of his evolution theory. This approach assumes that consciousness is a byproduct, a surface phenomenon, of the activity of the brain.

Different approaches have different advantages and disadvantages. Psychophysical interactionism has to give a reasonable answer to the crucial question “How?”. How can we explain that the non-physical consciousness affects and is influenced by physical processes? Psychophysical parallelism has to give a reasonable answer to the crucial question “Why?”. Why do we have consciousness when it does not matter to our lives or to evolution? Panpsychism has a special position among the parallelist theories in that it is possibly evades these questions. I will come back to this problem later. Epiphenomenalism has to give a reasonable answer to both “Why?” and “How?”. Why do we have consciousness if it does not matter in evolution? And how can physical processes affect the non-physical consciousness?

4 The Evolution Argument

As mentioned above, an important argument for Popper’s position is the evolution argument. An interactionist solution answers the question “Why?”. It is reasonable (but not necessary) to assume that consciousness has emerged during evolution and that organisms with consciousness had survival advantages over organisms without. This requires that conscious processes affect physical brain processes and vice versa.

Parallelist or epiphenomenalist solutions do not answer this question. In a parallelist theory, a description of consciousness is not required for a complete description of the world including humans and their actions; it seems that it is in principle possible to describe a person and her actions completely without assuming that she can experience anything, that she is conscious. But we know we are conscious and our consciousness is something extremely important for us! This is an old argument but it still weighs heavily. William James wrote in 1879:

> Consciousness is a manifested property of higher organisms, most evident in man; like all such characteristics it must have evolved; and it may only have been developed through natural selection; but if developed through natural selection it must have a use; and if it has a use it cannot be causally ineffective. (James 1879)

Popper also presented other arguments for an interactionist solution of the mind-brain problem. One refers to his three-world view on the universe. Since a world of objective theories, his World 3, exist and since we can use theories to modify the physical world, World 1, there must be a world that allows us to grip these World 3 elements. This is the world of subjective consciousness, World 2. And since we can work in world 1, the three worlds must be open to each other, i.e. interaction must be possible. Popper presented a number of arguments against parallelist theories beside the ones briefly mentioned above. Since I here focus on arguments related to the theory of evolution, I will highlight some of Popper's arguments against panpsychism, the parallelist theory that seems to be immune to evolution arguments.

5 Popper's Argument Against Panpsychism

Another reason for discussing panpsychism is that the interest in this theory has increased markedly in recent years (Smith 2008; Koch 2019). Originally the idea was introduced to eliminate the problem of how novelties emerge.

But Popper's point is that novelties do emerge. Solid ice becomes liquid water when temperature increases. It does not help us to understand the phase transition by introducing a concept of proto-liquidity in solid ice. It does not help us to understand the origin of consciousness to assume proto-consciousness in pre-biotic matter. Panpsychism seems to lack explanatory power. And even if stones would have some sort of proto-consciousness, we have to admit that animal consciousness seems so much richer than proto-consciousness of a stone that the difference between them becomes so great that in practice it would be difficult to distinguish a panpsychist explanation from explanations that assume that consciousness emerges during evolution.

Popper raised another argument against panpsychism that I find especially interesting (Popper and Eccles 1977, 69–71). It relates to quantum physics and concerns the question whether atoms and elementary particles have some form of memory. There are many reasons to assume that consciousness involves some form—albeit short-lived—of memory mechanism. It is hard to imagine conscious experience that does not include some kind of continuity over time. Popper makes a *Gedanken*-experiment to prove his point; atoms and elementary particles should, according to panpsychism, have memory-like properties. But many contemporary physicists emphatically stress that atoms and elementary particles lack memory. Two radioactive atoms of the same isotope have the same propensity for decay irrespective of their history. However, this might not be entirely uncontroversial. It has also been argued that quantum physics is compatible with a metaphysics of individual objects, but that such objects are indistinguishable in a sense, which leads to the violation of Leibniz's famous Principle of the Identity of Indiscernibles. In summary, the prevailing view today is that fundamental particles of physics cannot be regarded as individual objects, thus making panpsychism an unlikely solution of the consciousness problem.

6 Problems with the Interactionist Position

The main problem with the interactionist view is that it seems incompatible with the conservation laws of physics (Wilson 1999; Clarke 2014). Popper responded repeatedly and specifically to these arguments. But in general, he did not seem especially worried about this criticism. This was of course due to his hypothesis of a three-world universe and his central thesis that present-day physics is fundamentally incomplete; that the universe is open. Nevertheless, he gave specified arguments to his critics. His most often repeated argument was that the first law may only be statistically valid (Popper and Eccles 1977, dialogues X and XII). This is an idea first suggested by Schrödinger (1952).

> One possibility that would suit us extremely well would be that the law of the conservation of energy would turn out to be valid only statistically. If this is the case, it might be that we have to wait for a physical fluctuation of energy before world 2 can act on world 1, and the time-span in which we prepare for the "free-will movement of the finger" may easily be long enough to allow for such fluctuations to occur.

Another argument was that there might exist 'purely mental forms of energy, convertible into electrochemical forms' (Popper 1984, 21).

A further argument was that, according to some interpretations of de Broglie's particle-wave theory, 'there seem to be empty pilot waves that can interfere with non-empty (energy-piloting particles an energy-carrying) waves', and this would suggest 'the possibility of non-energetic influences upon energetic processes' (Popper 1984, 21–22).

7 Popper's New Theory of Mind

As mentioned above, Ingemar Lindahl and I had the opportunity on two occasions to interview Popper about what he called his new theory of mind (Popper et al. 1993, 1994, 2010). In these he pointed to the similarity between mind and forces, by characterizing mind as being: (i) located (ii) unextended (iii) incorporeal (iv) capable of acting on bodies (v) dependent upon body (vi) capable of being influenced by bodies (vii) intensities, and (viii) extended through a span of time.

> Most people would say, I think, if one tells them that something with all these properties exists, that it cannot be true. Especially, most materialists would say so, and most physicalists. Now, I say things of this kind do exist, and we all know it. So, what are these things? These things are forces. For example, electrical forces. Electrical and magnetic forces have all these properties. (Popper et al. 1993)

This similarity between mind and forces is not a new discovery. As Popper pointed out in "The Self and its Brain" (1977), the analogy between mind and forces has been used before. Hobbes and Leibniz identified a certain part of mind with a physical force. Gilbert in De Magnete "had compared the interaction between magnetic force and a loadstone to that between soul and body". Both Thomas Reid and Maine de

Biran emphasized our experience of the mind (the will) acting on our body and producing effects in the material world as the source of our universal notion of force. But Popper went further. He specified the type of force field he was thinking of, an electromagnetic field:

> I wish to propose here as a hypothesis that the complicated electro–magnetic wave fields which, as we know, are part of the physiology of our brains, represent the unconscious parts of our minds, and that the conscious mind — our conscious mental intensities, our conscious experiences — are capable of interacting with these unconscious physical force fields, especially when problems need to be solved that need what we call 'attention'". This admittedly vague working hypothesis seems to me as a small yet significant progress within a so far hopelessly difficult part of physiology. (Popper et al. 1993)

Popper seemed to view the "unconscious parts of our minds" as synonymous with "physical force fields". And he seemed to view the electromagnetic field (the unconscious) as an intermediate link between the conscious mind and the neuronal activity. This is perhaps the most central thesis in his new hypothesis of mind and I will come back to it below. But first I will mention another thought-provoking observation Popper took up in the interview, the relative autonomy of forces.

8 The Autonomy of Forces

> To what extent, if any, can a concept of physical force account for the apparent autonomy of mind? … We tend to think of forces as something attached to bodies, and not as something that can obtain autonomy…. The fundamental question is: "How can these forces, which are set up in the brain, continue themselves, so to speak, and continue to have a kind of identity which is even able to initiate in its turn biochemical processes in the brain?" (Popper et al. 1993)

Popper seemed to think that very few physicists had seen this autonomy of forces as a problem. One of the few was the Swedish physicist Hannes Alfvén (Nobel laureate 1970).

> In Alfvén's cosmology, forces, mainly electrical forces, but of great complexity, are living in the cosmos everywhere. Like the forces which creates the northern lights. To put it in another way, apart from the stars, which, of course, he admits exist, there exists a semi-matter, like electrons, without density of distribution, but with forces holding them together; electrical forces. The forces are partly the effect of the electrons. This phenomenon is as, let us say, weeds drifting in the sea. One does not know exactly why the electrons are together. They are not attracting each other. Somehow the electrons modify the situation.
>
> I mean, electrons, in any case, are not what we usually call matter. The electrons held together, forming curtains, like the real curtain-like arrangement of the northern lights are here repeated all through the cosmos.
>
> I do not think that Alfvén would claim to know all about forces. But he would say, yes forces are, in a sense, almost independent. Nothing is independent, but the forces are as independent as matter, in our space. That is roughly the situation. (Popper et al. 1994)

9 An Interpretation of Popper's Theory

In 1994 Ingemar Lindahl and I published an interpretation of Popper's new theory of mind. We took as point of departure his identification of unconscious mind with certain electromagnetic fields, and suggested a three-level scheme, depicted in Fig. 1 below (Lindahl and Århem 1994). It shows the relation between consciousness, or the conscious part of mind, the electromagnetic fields of the brain and what we called the action potential pattern of the brain. Perhaps a better label of the lowest level would have been electric current patterns of the brain; this perhaps would be more in line with at least one strand of Popper's realistic interpretation of the physical world (Popper 1982a); a world containing particles and fields (currents are moving electrons). The figure also shows the relations between worlds 1 and 2 as well as the relation between mind and brain.

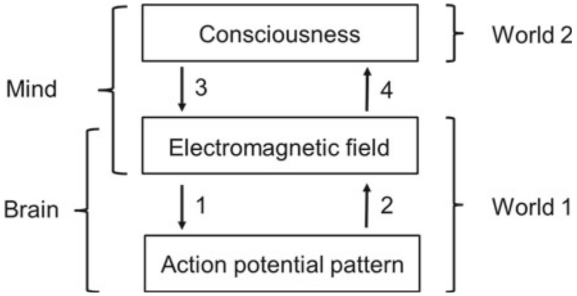
According to this interpretation, there are two levels of interaction: the first between the currents associated with certain spatio-temporal patterns of action potentials and specific electromagnetic fields (the relations 1 and 2); the other between the electromagnetic fields/the unconscious, and the conscious mind (the relations 3 and 4).

The introduction of an intermediate link, the electromagnetic field/the unconscious, that have properties in common both with conscious mind (the eight properties) and with the spatio-temporal pattern of action potentials (the membership of world 1), may make it somewhat easier to conceive of an interaction between consciousness and the brain, two very different entities in themselves.

Relation 2 appears to be the least problematic of the four in the figure. This relation may in principle be studied within classical electrodynamics.

Relation 1 may seem more difficult to accept. In order to excite a resting, inactive neuron, it is necessary to change the membrane potential by 20 mV (Hille 2001). Simple calculations show that under the most favourable conditions an electric field of at least 0.5 V/cm would be necessary. However, the electric field around a nerve cell, induced by a normal impulse activity, is many times weaker due to the low resistance of the extracellular part of the local circuit. Thus, under these circumstances, an electromagnetic field effect on the brain seems highly unlikely. However, according to Popper (Popper and Eccles 1977, dialogue X), the electromagnetic field is not

Fig. 1 (After Lindahl and Århem 1994)



expected to trigger inactive neurons, but to sculpture ongoing neuronal activity; to affect neurons in constant spontaneous activity.

> Thus, what I am here suggesting is that we might conceive of the openness of World 1 to World 2 somewhat on the lines of the impact of selection pressures on mutations. The mutations themselves can be considered as quantum effects; as fluctuations. Such fluctuations may occur, for example, in the brain. In the brain there may at first arise purely probabilistic or chaotic changes, and some of these fluctu-ations may be purposefully selected in the light of World 3 in a way similar to that in which natural selection quasi-purposefully selects mutations. […]. (Popper and Eccles 1977, 540)

In a next step, Popper even suggested that the so called all-or-nothing principle of nerve cell firing may be the mechanism of allowing microscopic effects to be macroscopic:

> The all–or–nothing principle of the firing of nerves may indeed be interpreted as a mechanism which would allow arbitrarily small fluctua-tions to have macroscopic effects…. The action of the mind on the brain may consist in allowing certain fluctuations to lead to the firing of neurones while others would merely lead to a slight rise in the temperature of the brain. (Popper and Eccles 1977, 541)

In referring to "quantum effects" Popper seemed to refer to truly random (i.e. not only in practice difficult to predict) neuronal activity. That can mean either indeterminate quantum effects, described by the Heisenberg principle in some form, or it can mean, more controversially, macroscopic indeterminate effects, described by Popper's propensity theory (Popper 1982a, 1990).

Thus, according to Popper's new theory of mind in our interpretation, it is mind effects on the electro-magnetic field that modulate the nerve cell firing. It is not direct effects on critical nerve cell structures that modulates their firing. These structures are the ion channels, membrane proteins that selectively allows metal ions to flow through the membrane. Ion channels will be discussed in more detail Sect. 12.

Two observations in my own lab may have some bearing on this issue. One was that opening of a single channel may cause certain neurons to fire action potentials (Johansson and Århem 1994). Normally thousands of channels are necessary for a neuron to fire an action potential. For a single channel current to induce a sufficient potential change, either the current or the membrane resistance must be unusually large (due to Ohm's law). We succeeded to demonstrate such an effect in certain brain cells. These studies thus suggest that an extremely small effect may be amplified to trigger all-or-nothing action potentials in cells of the brain, and consequently to trigger activity in circuits and larger brain networks.

The second observation was that there seem to be "real" thresholds and "pseudo"-thresholds for triggering action potentials in neurons. A mathematical analysis of the excitability of different neuron models showed that some had a discontinuous current voltage-relation (a real threshold), and some had a continuous, albeit very steep current-voltage relation (a pseudo-threshold) (Århem and Blomberg 2007; Zeberg et al. 2010, 2015). The reason could be traced back to different bifurcation properties (Izhikevich 2007). This means that there may be neurons that are easier triggered than others, in theory by an infinitely small voltage change.

Popper's hypothesis is of course highly speculative. But such speculations seem necessary in order to get us somewhere in coping with this extreme problem. Popper's attempt to correlate mental processes with electromagnetic fields is an attempt to formulate a realistic hypothesis. Of course, Popper is not alone in discussing consciousness in field terms (for a summary, see Jones 2013). Benjamin Libet even offered an idea that is experimentally testable (Libet 1994, 1997). As stated previously:

> Our main conclusion is that Popper's hypothesis of consciousness interacting with neural activity through an electromagnetic field is a thought-provoking suggestion worth closer examination; and that his theory of mind as a whole is possibly the most promising proposal yet made for a future explanation of the survival and development of consciousness. (Lindahl and Århem 1994)

10 Beck's Interpretation of Popper's Theory

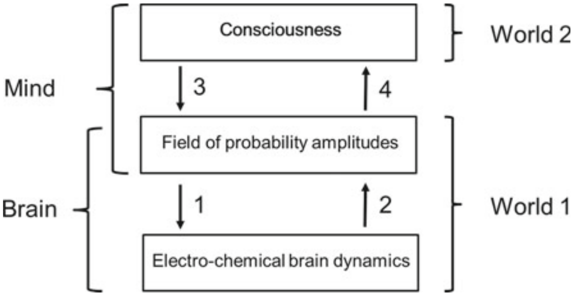
Friedrich Beck (1996) commented on our interpretation above (Lindahl and Århem 1994). His main criticism was that the direct relation between consciousness and the electromagnetic fields is in conflict with the conservation law of energy. To avoid this he suggested a direct relation between consciousness and probability fields of quantum mechanics, as depicted in Fig. 2 below.

The suggestion shows clear similarities with Popper's hypothesis. It also uses the argument of a family resemblance between fields and consciousness. The physicist Henry Margenau has previously discussed the resemblance between consciousness and quantum probability fields:

> The mind may be regarded as a field in the accepted physical sense of the term, but it is a non-material field, its closest analogue is perhaps a probability field …. nor is it required to contain energy in order to account for all the known phenomena in which mind interacts with the brain. (Margenau 1984).

But does this suggestion by Beck evade conflicts with other principles of physics? I do not think so. Quantum probability fields are stochastic and do not allow modulation outside the fixed values of the statistical parameters in present-day physics. So the

Fig. 2 (After Beck 1996)



question is how much more reasonable the probability field solution is than the electromagnetic field solution. This will be discussed in the next section.

11 Mind Affecting Probability Fields: The Microsite Hypothesis

The proposal by Beck above draws on quantum mechanical principles. He and Eccles have developed a hypothesis assuming that mental effects modifies quantum mechanical probability fields involved in synaptic transmission. They called this hypothesis the microsite hypothesis. Given Popper's long-standing interest in quantum mechanics (Popper 1982a) and his close friendship with Eccles, it may seem remarkable that he did not comment more on this hypothesis. Rather, in Ingemar Lindahl's and my discussion with him (Popper et al. 1994), he seemed rather uninterested in the quantum mechanical aspects of the mind-brain problem, illustrated by his remark "It is one of the hundred ways".

This seeming lack of interest may be related to his well-known criticism of the mainstream interpretations of quantum mechanics, as developed in "Quantum Theory and the Schism in Physics" (1982a). But again, considering his three-world view with its interacting worlds, it does not seem necessary to evade conservation laws of present-day physics with present-day physical principles. As stated in "The Open Universe", the universe is fundamentally open (Popper 1982b).

As mentioned above, Beck and Eccles developed the microsite hypothesis in 1992. Eccles had searched for quantum mechanical solutions for a long time (Eccles 1987, 1992), but did not find them successful until he established the collaboration with Beck. In his early attempts he assumed that mental events directly affected the vesicles containing neurotransmitters. With Beck, however, he assumed that the specific target was presynaptic grids (a structure within the synaptic bouton attached to apical dendrites of pyramidal cells, first described by (Akert et al. 1975)). By these quantum mechanical assumptions they hoped to avoid breaking the first law of thermodynamics. A reason, as mentioned above, not important for Popper.

The microsite hypothesis of Beck and Eccles assumes the existence of critical quasiparticles, the mass of which being in the range of hydrogen atoms, and therefore within the quantum mechanical regime, rather than in the classical thermodynamic regime. Mind is assumed to affect these quasiparticles located in the presynaptic grid by modifying their probability fields. This triggers a release of transmitter by movement of the quasiparticle through an energy barrier via the process of quantum tunneling, meaning that for a particle of a certain energy there is a finite probability of penetration through a barrier even when the particle energy is less than the barrier energy, something that is impossible in the thermodynamic regime. The probability of a particle penetrating the barrier is given by the transmission coefficient, which is a function of the shape and amplitude of the energy barrier, the mass of the particle and other factors. Beck and Eccles estimated the transmission coefficient to be between

0.4 and 4%, meaning that the probability of a synapse releasing transmitter substance at a nerve impulse is 25%. (The figures will be used in an illustration below).

The microsite hypothesis was not received uncritically by the neuroscience community, nor by the philosophy-of-mind community (see Wilson 1999; Clarke 2014). The main problem is of course that even if the conservation of energy is preserved in the hypothesis, it introduces an effect (mind modifying quantum probability fields) not belonging to present-day physics. Furthermore, the understanding of the process of synaptic release has evolved since the publication of the microsite hypothesis and seems not to support the processes Beck and Eccles assumed to be at hand.

An experimental finding seemingly incompatible with the hypothesis is the fact that processes in the quantum regime should be temperature independent, but transmitter release is temperature dependent. Such problems can be accounted for by making ad hoc assumptions, but this, of course, does not strengthen the hypothesis. In summary, the microsite hypothesis in its present form seems unlikely as explanation of the interaction between mind and brain.

12 Mind Affecting Probability Fields: The Ion Channel Hypothesis

Popper's new hypothesis assumed that consciousness interacts with brain activity via an intermediate level, the electromagnetic field level. Beck suggested interaction via quantum probability fields. Both hypotheses use the family resemblance argument. Let us explore the idea that consciousness interacts with the electric activity of nerve cells via modifying quantum probability fields associated with critical structures of ion channels. The impulse initiation in nerve cells seems today better understood in terms of molecular details than synaptic impulse transmission, which forms the basis of the microsite hypothesis. This gives us possibility to explore the role of quantum mechanics in brain function in more detail than when studying corresponding problem in synaptic processes.

Since the studies by Erwin Neher and Bert Sakmann in the early 80's, we know that the current through the membranes pass through pores of special proteins, called ion channels. With extremely sensitive technology, we can now directly study the activity of single ion channel molecules in real time. Through such studies in combination with molecular biology, we now know a lot about these ion channels (see Hille 2001). There are a wide variety of different types (143 species in the class of voltage-gated channels in humans forming the human channelome); there are channels selectively permeant to sodium, potassium and calcium ions; there are channels activated by the electrical voltage across the membrane and there are channels that are activated by specific molecules. Each nerve cell has its particular palette of ion channels depending on its function. For the passage of the nerve impulse along the nerve fibers, voltage-activated sodium and potassium channels play a major role.

The ion channels have a long evolutionary history. The same voltage-activated potassium channels that contribute to the nerve impulses in humans are found in some single-celled organisms that are present some 1400 million years ago. Sodium channels that play the leading role in human impulse conduction are found in the evolutionary early cnidarians (i.e. jellyfish, hydras and corals), perhaps 700 million years ago (Hille 2001). Consciousness apparently does not depend on specific molecules, specifically human consciousness does not depend on specific human molecules. There does not seem to be any specific human ion channels. The same molecules are found far down the phylogenetic chain.

For similar reasons, there does not seem to be specific neurons in species that can be assumed to be conscious. There are no specific human nerve cells. It seems likely that the emergence of a consciousness has to do with the organization and the processes of the nervous system, and that the emergence of a specific human consciousness has to do with the specific organization of the human brain.

Ion channels are membrane-bound proteins with a central ion-permeant pore. The voltage-sensitive component of voltage gated channels is an electrically charged helical structure, called the S4 segment. The details in the opening sequence are still debated, but a widely accepted idea is that the S4 moves outwards in a screw-like fashion, initiating a sequence of events that open the pore.

The mathematics of nerve excitability and the functioning of ion channels is relatively well developed. It is based on the ideas presented Allan Hodgkin and Andrew Huxley in 1952, extended with new mathematical tools such as bifurcation theory (Izhikevich 2007), explaining the firing patterns of different types of neurons, and the transition state theory of Henry Eyring and Michael Polanyi, explaining the rate of opening.

Let us assume mental events modify quantum probability field associated with the S4 segment. The S4 segment has a 200-fold higher mass than the quasiparticle postulated by Beck and Eccles in their microsite hypothesis, suggesting that in the channel opening case we are operating in the borderland between a quantum regime and a thermodynamic regime. What does this mean for the opening of the channel? Are quantum mechanical principles irrelevant for describing nerve firing? Probably not.

According to the transition rate theory, the rate constant for the opening is a complex function of the tunneling probability (i.e. the transmission coefficient), which in turn is a function of the mass of the critical structure and the barrier height. The barrier height depends on the membrane voltage and consequently on the specific spatial and temporal conditions for the neuron studied. To analyze this thoroughly we need massive computer power. However, for the present purpose we can use a simpler approach.

It does not seem unreasonable to assume that the transmission coefficient in some situations is within the range of the transmission coefficient estimated by Beck and Eccles, i.e. between 0.4 and 4%. Thus, assuming a transmission coefficient of 0.4%, conventional computer simulations of nerve cell activity using the Hodgkin-Huxley formalism show that the firing frequency differs measurably between that of a nerve cell model assuming tunneling and a nerve cell model without such an assumption, suggesting that quantum mechanical principles may play a role at a macroscopic level.

However, it does not show that mind acts in this way. Such a theory must explain how the suggested marginal effect is filtered out from thermodynamically initiated firing, how the thousands of channels are coordinated, and why certain cortical cells are affected. A plethora of biophysical, systems-biological and evolutionary questions remain to be answered.

From a Popperian point of view, I do not think performing detailed quantum mechanical calculations is the way to go. The main reason for this approach was the wish to evade the energy conservation laws. And as mentioned above, this was not an important issue for Popper.

13 Concluding Remark

In conclusion, no solution of the mind-brain problem seems to be in sight, not within present-day science boundaries. Nevertheless, the hypothesis discussed in the present chapter seems to offer some hope. In Popper's words (Popper et al. 1994):

> This admittedly vague working hypothesis seems to me as a small yet significant progress within a so far hopelessly difficult part of physiology.

Our interpretation of this hypothesis assumes an intermediate level between conscious mind and brain, and that this mediating stratum consists of specific electromagnetic fields in the brain.

It should be noted that mind-brain theories assuming a role of electromagnetic fields are not new. However, these theories often identify mind with an electromagnetic field (see Jones 2013; McFadden 2013). Consequently, they belong to the parallelist camp and can be criticized accordingly. In addition, they seem to fall prey to Leibnitz' law of the identity of indiscernibles.

As a final remark, I would like to comment on Popper's criticism of instrumentalist attitudes (Popper 1963). Many areas remain to be explored to get us closer to an understanding of the mind-brain relation. Specific questions related to Popper's hypothesis are: Which brain cells or groups of brain cells are selected by mind, what physical criteria characterize these cells, and why are they selected? How are the microscopic events amplified to macroscopic events? How are the induced microscopic fluctuations isolated and shielded from thermodynamic noise?

These questions are mainly neuroscience questions. But to understand mind-brain issues it is imperative to understand underlying neuroscience issues. And to do that, Popper's realist, anti-instrumentalist approach seems essential (Popper 1963); it inspires us to transform suggestions into realistic hypotheses that can be experimentally tested. All ways to approach the mind-brain problem must be continually confronted with neurophysiological findings. This is not always done. Instrumentalist black-box attitudes are unfortunately not uncommon in large areas of mind-brain studies; e.g. in studies based on computational or cognitive neuroscience. I am afraid such attitudes may hamper our attempts to make any progress within, to use the words of Popper again (Popper et al. 1993), “a so far hopelessly difficult part of physiology”.

References

Akert, K., Peper, K., Sandri, C.: Structural organisation of motor end plate and central synapses. In: Waser, P.G. (ed.) Cholinergic Mechanisms, pp. 43–57. Raven, New York (1975)

Århem, P., Blomberg, C.: Ion channel density and threshold dynamics of repetitive firing in a cortical neuron model. Biosystems **89**(1–3), 117–125 (2007)

Beck, F.: Mind-brain interaction: comments on an article by B. I. B. Lindahl & P. Arhem. J. theor. Biol. **180**, 87–89 (1996)

Beck, F., Eccles, J.C.: Quantum aspects of brain activity and the role of consciousness. Proc. Natl. Acad. Sci. **89**, 11357–11361 (1992)

Bloch, M.: In and out of each other’s bodies: theories of mind, evolution, truth, and the nature of the social. Paradigm, Boulder (2013)

Clarke, P.G.H.: Neuroscience, quantum indeterminism and the Cartesian soul. Brain Cogn. **84**(1), 109–117 (2014)

Crick, F., Koch, C.: Towards a neurobiological theory of consciousness. Semin. Neurosci. **2**, 263–275 (1990)

Dennett, D.C.: Consciousness explained. Little, Brown and Company, Boston (1991)

Descola, P.: Beyond nature and culture. University of Chicago Press, Chicago (2013)

Eccles, J.C.: Do mental events cause neural events analogously to the probability fields of quantum mechanics? Proc. R. Soc. Lond. B **227**, 411–428 (1987)

Eccles, J.C.: The evolution of consciousness. Proc. Natl. Acad. Sci. **8**, 7320–7324 (1992)

Edelman, G.M.: Bright air, brilliant fire: on the matter of the mind. Allen Lane the Penguin Press, Harmonsworth (1992)

Feigl, H.: The “mental” and the “physical”. University of Minnesota, Minneapolis (1967)

Hille, B.: Ionic channels of excitable membranes, 3rd edn. Sinauer Associates Inc, Sunderland (2001)

Izhikevich, E.M.: Dynamical systems in neuroscience: the geometry of excitability and bursting. MIT Press, Cambridge (2007)

James, W.: Are we automata? Mind **4**, 1–22 (1879)

Johansson, S., Århem, P.: Single-channel currents trigger action potentials in small cultured hippocampal neurons. Proc. Natl. Acad. Sci. **91**, 1761–1765 (1994)

Jones, M.W.: Electromagnetic-field theories of mind. J. Conscious. Stud. **20**(11–12), 124–149 (2013)

Koch, C.: The feeling of life itself—why consciousness is widespread but can’t be computed. The MIT Press, Cambridge (2019)

Libet, B.: A testable field theory of mind–brain interaction. J. Conscious. Stud. **1**, 119–126 (1994)

Libet, B.: Conscious mind as a force field: a reply to Lindahl & Århem. J. Theor. Biol. **185**, 137–138 (1997)

Lindahl, B.I.B., Århem, P.: Mind as a force field: comments on a new interactionistic hypothesis. J. Theor. Biol. **171**, 111–122 (1994)

Margenau, H.: The miracle of existence. Ox Bow Press, Woodbridge (1984)

McFadden, J.: The CEMI field theory: closing the loop. J. Conscious. Stud. **20**(1–2), 153–168 (2013)

Niemann, H.-J.: Karl Popper and the two new secrets of life. Mohr Siebeck, Tübingen (2014)

Popper, K.R.: Language and the body–mind problem: a restatement of interactionism. In: Proceedings of the XIth International Congress of Philosophy, vol. 7, 101–107. North-Holland, Amsterdam (1953)
Popper, K.R.: A note on the body–mind problem. Analysis **15**, 131–135 (1955)
Popper, K.R.: Conjectures and refutations: the growth of scientific knowledge. Routledge & Kegan Paul, London (1963)
Popper, K.R.: Objective knowledge: an evolutionary approach. Clarendon, Oxford (1972)
Popper, K.R.: Indeterminism is not enough. Encounter **40**, 20–26 (1973)
Popper, K.R.: Unended quest: an intellectual autobiography. Fontana/Collins, London (1976)
Popper, K.R.: Some remarks on panpsychism and epiphenomenalism. Dialectica **31**(1–2), 177–186 (1977)
Popper, K.R.: Natural selection and the emergence of mind. Dialectica **32**, 339–355 (1978)
Popper, K.R.: Quantum theory and the schism in physics. Unwin Hyman, London (1982a)
Popper, K.R.: The open universe: an argument for indeterminism. Hutchinson, London (1982b)
Popper, K.R.: Critical remarks on the knowledge of lower and higher organisms the so-called sensory motor systems. In: Creutzfeldt, O., Schmidt, R.F., Willis, W.O. (eds.) Sensory-Motor Integration in the Nervous System, pp. 19–31. Springer-Verlag, Berlin (1984)
Popper, K.R.: A world of propensities. Thoemmes, Bristol (1990)
Popper, P.R.: Knowledge and the body-mind problem. Routledge, London (1994)
Popper, K.R., Eccles, J.C.: The self and its brain: an argument for interactionism. Springer-Verlag, Berlin (1977)
Popper, K.R., Lindahl, B.I.B., Århem, P.: A discussion of the mind–brain problem. Theor. Med. **14**, 167–180 (1993)
Popper, K.R., Lindahl, B.I.B., Århem, P.: A discussion of the mind–brain problem. (Manuscript) (1994)
Popper, K.R., Lindahl, B.I.B., Århem, P.: Die Beziehung zwischen Bewusstsein und Gehirn: Disskussion einer interaktionistischen Hypothese. Aufklärung und Kritik **4**, 9–21 (2010)
Ryle, G.: The concept of mind. Hutchinson, London (1949)
Searle, J.R.: Mind: a brief introduction. Oxford University Press, New York (2004)
Schrödinger, E.: Are there quantum jumps? Part I-II. Br. J. Philos. Sci. **3**(109–123), 233–242 (1952)
Smith, C.U.M.: Synapses, quantum theory and panpsychism. NeuroQuantology **6**(2), 164–174 (2008)
Solecki, R.S.: Shanidar. Knopf, New York (1971)
Wilson, D.L.: Mind-brain interaction and violation of physical laws. J. Conscious. Stud. **6**(8–9), 185–200 (1999)
Zeberg, H., Blomberg, C., Århem, P.: Ion channel density regulates switches between regular and fast spiking in soma but not in axons. PLoS. Comput. Biol. **6**(4), 22 (2010)
Zeberg, H., Robinson, H.P., Århem, P.: Density of voltage-gated potassium channels is a bifurcation parameter in pyramidal neurons. J. Neurophysiol. **113**(2), 537–549 (2015)

Peter Århem is professor in neurophysiology at Karolinska Institutet, Stockholm, Sweden. He is studying the function of nerve cells and related mathematical and philosophical issues. He has contributed as author/editor to several books such as "Matter Matters" (1997) and "Consciousness Transitions" (2008). Together with Ingemar Lindahl, he has published two interviews with Karl Popper on the mind-brain problem. E-mail: peter.arhem@ki.se

Karl Popper on the Evolution of Consciousness

Manjari Chakrabarty

1 Introduction

Some scholars (e.g., Skoyles 1992) feel that Popper's work has been widely acknowledged by the scientists but has had little impact on professional philosophers. Some others (e.g., Lindahl 1992) are of the opinion that Popper's influence on both the philosophers and the scientists has been considerable, or to borrow Bondi's (1992, 363) words, "…Popper's influence shines through." However, it wouldn't be incorrect to say that Popper's teachings and views on the evolution of consciousness (or minds) and on the consciousness-brain or minds-brains interactions have received comparatively little scholarly attention from mainstream philosophers of mind. His name does not appear in many introductory books and edited volumes on philosophy of mind (see, e.g., Churchland 1984; Lowe 2000a; Heil 1998, 2004).

Popper has been actively interested in the key issues related to the philosophical theories of mind and its relation to the brain, for many years. His first papers on the subject now reprinted as Chaps. "Language and the Body-Mind Problem" and "A Note on the Body-Mind Problem" of *Conjectures and Refutations* (1963) were published in the early 1950s. A more fully developed interactionist hypothesis—intertwined with his conjecture of three worlds—has then appeared in his 1977 publication *The Self and Its Brain*, written in three parts with John Eccles. Popper has proposed this (non-dualist) interactionist hypothesis mainly to address the classical problem of interaction, namely, how two very different entities such as minds and brains can possibly interact, from an evolutionary (biological) perspective. Ever since its first explicit presentation the said hypothesis has been disapproved by several philosophers (see e.g., Dennett 1979; Rooijen 1987; Settle 1989). One possible reason behind the philosophers' strong resistance to Popper's (non-dualist) interactionist hypothesis could be the growing dominance of some form

---

M. Chakrabarty (✉)
Visva Bharati (Central University), Santiniketan, India
e-mail: manjari.chakraborty@visva-bharati.ac.in

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_16

of materialism or physicalism[1] in philosophical circles since late 1970s along with its tendency to reject all conceivable versions of consciousness-brain interactions. Besides, Popper and his co-author Eccles are often misrepresented by prominent philosophers (e.g., Lowe 2000b, 575) as ‘interactionist dualists’. In such an intellectual climate it is hardly surprising that Popper’s (non-dualist) interactionism remains an unpopular, and comparatively less well-examined position in the field.

In the early 1990s a ‘new theory of mind’ (Popper et al. 1993, 168), partly based on his earlier interactionist hypothesis, has been introduced by Popper. This new theory, presently known as the ‘mental force field’ hypothesis, characterizes minds as having important similarities with recognized physical forces (Popper et al. 1993). Although this novel hypothesis of Popper has been further extended by two Swedish philosophers B.I.B. Lindhal and P. Århem (e.g., Lindahl and Århem 1994) and has also been examined by some well-known neurophysiologists (e.g., Libet 1996; Jones 2013), it is yet to found a serious place in contemporary philosophical investigations into the nature of consciousness. Probably because of Popper’s unconventional[2] portrayal of consciousness as an emergent, biological yet force-like phenomenon and his silence on the subjective aspect of consciousness (see e.g., Nagel 1974), his mental force field hypothesis doesn’t interest many who attempt to define consciousness in terms of its characteristic privacy and qualitative nature.

In the present chapter, consisting of three sections, it is argued that a critical review of Popper’s (non-dualist) interactionist hypothesis (supplemented by his mental force field hypothesis) is both urgent and necessary for a more comprehensive understanding of the evolution of consciousness and its dynamic interactions with the brain. The reasons assumed to be crucial for substantiating this core argument are stated below. First, a closer scrutiny of the hypotheses of Popper reveals how consciousness-brain interactions may plausibly be explained (without violating the laws of classical physics) and thereby a challenge may be posed to their apparently irrefutable rival, physicalism (the characteristic principle of which is the closedness of the physical world). The next two sections deliberate on the above reason.

The second reason concerns the philosophers’ utter neglect of the burgeoning archaeological research on the evolution of (hominin) mental faculties, over the past decades. In an exchange of views (published in the journal *Mind and Language*) two of the most renowned philosophers of mind, Dennett (1996) and Fodor (1996), disagreeing fundamentally on the issues regarding the evolution of mental faculties, agree with each other on the point that these issues cannot be addressed ‘until the data is in’. If only these philosophers took a moment to examine the experimental-archaeological literature, noted archaeologist Steven Mithen (1998, 5) argues, they could have seen that a huge amount of the relevant data is not only ‘in’ but has already

---

[1] Despite having very different histories the terms ‘materialism’ and ‘physicalism’ are used interchangeably here.

[2] Popper’s hypothesis seems unconventional because it neither promotes the ‘mind-as-computer’ view nor the ‘mind-as-brain’ view (Jones 2010).

been subject to immense archaeological interpretation and analysis. Popper's (non-dualist) interactionist hypothesis is quite a unique philosophical account that focuses on the evolutionary (pre) history of minds or of the different levels of consciousness. Expressing his discomfort regarding the common, injudicious use of the phrase 'the conscious mind' (Popper 1994, 111), Popper has focused instead on the many different levels of consciousness and on their biological significance.

This underlying belief of Popper in the existence of different levels of consciousness points towards the question of the evolution of consciousness in a world hitherto purely physical in its attributes. Interestingly, the decades-old Popperian speculations about the evolution of minds—both prehuman and human—look strongly convergent with current experimental-archaeological research. The final section of the present chapter intends to bring to light the convergence of Popper's philosophical views and recent archaeological explorations that has gone largely unnoticed in the relevant literature.

On basis of the reasons stated above the present chapter concludes with the argument that Popper's (non-dualist) interactionist hypothesis (supported by his mental force field hypothesis)—is neither an explanation "… we can perhaps afford to ignore… in philosophical discussions related to causal closure principles and emergentism…" (Lowe 2000b, 575) nor one that "…fails to make serious contact with the best theoretical work of recent years…" (Dennett 1979, 91)—but is a serious contender in the philosophical battlefield that deserves more critical attention from the mainstream philosophers of mind.

2 In What Sense Is Consciousness Distinct from the Brain?

Two questions are most critical for Popper. First, in what sense is consciousness distinct from the activities and states of the brain? Second, how does consciousness—seen as distinct from the brain—causally interact with the brain? While Popper's (non-dualist) interactionist hypothesis puts more emphasis on the first question, his later (mental force field) hypothesis addresses the second question more directly. The present section attends to the first question and the second question will be considered in the following section.

Popper has explicitly stated that his 'tentative' but 'testable' (Popper 1994, 105) interactionist hypothesis complicates the process of explanation by construing consciousness as distinct from the brain. Physicalism, in contrast, seems 'intrinsically convincing' (Popper and Eccles 1977, 51) and hard to decisively refute because by denying the existence of consciousness as distinct from the brain it simplifies the explanatory task and wipes out a number of difficult problems (Popper 1994, 105). Nevertheless, physicalism is unacceptable to Popper because it dogmatically explains away one of philosophy's greatest riddles instead of seriously investigating it (Popper and Eccles 1977, 53).

Popper's answer to the question—in what sense is consciousness distinct from the activities and states of the brain—is implicit in his theory of three worlds (Popper and Eccles 1977; Popper 1979). Reality, for Popper (Popper and Eccles 1977; Popper 1979), is a tripartite phenomenon composed of an interacting triad of evolutionary levels, namely, world 1, world 2 and world 3. Each of these worlds or evolutionary levels is an irreducibly emergent[3] phenomenon and all three causally interact with one another. World 1 is the physical world of matter and energy including stars, planetary systems, physical bodies, forces, fields of force, living organisms and all organismic physical, chemical, biological processes. At some time in the distant past, prior to approximately 3.5 billion years, by processes not yet completely understood, life in the form of unicellular, micro-organisms (protobacteria) emerged from non-living matter. Life became complex progressively, as plants and animals of myriad forms and sizes evolved and interacted with fertile ecosystems.

During the evolution of life (in the form of microorganisms) on earth, some organisms became conscious (in a certain sense) and advantageously adapted to the contingencies of their environments. A new, i.e., qualitatively different realm of (primitive) conscious states and its attendant subjective experiences (e.g., feelings of pain, pleasure)—called world 2—emerged out of world 1 at a particular stage of evolution, though we cannot pinpoint that event. What makes world 2 as real as the physical world 1 is the fact that it causally interacts with the latter. The ability of causally acting on (ordinary-sized) physical bodies and being influenced by them is the core (and sufficient) criterion of reality for Popper (Popper and Eccles 1977, 9–11).

Out of the dynamic interplay of world 2 and world 1, probably at the stage corresponding to the appearance of some initial form of linguistic behaviour, emerges world 3—a distinct, extra-somatic, objective realm of human (also hominin) creation. Theories, propositions, the abstract yet objective contents of scientific, mathematical, or poetic thoughts, problem-situations and critical arguments have been described by Popper (1979) as the most fertile world 3 objects. Nevertheless, this world 3 or the ‘world of culture’ (Popper 1982, 54) includes myths, fairy tales, ethical values, social institutions, paintings, sculptures, and ‘feats of engineering’ such as, tools, machines, and scientific instruments (Popper 1979, 2). Manifestly, world 3 products are not genetically coded but can be modified, criticized, reinterpreted, and maintained (largely) by human beings. That world 2 is essentially instrumental in generating world 3 does not, however, call the reality of world 3 into question because world 3 can also act on world 1 (via world 2) and be acted upon (Popper and Eccles 1977).[4]

In the light of Popper’s tripartite account of reality comprising of three distinct but ‘somewhat overlapping’ (Popper and Eccles 1977, 48) and interacting worlds or evolutionary levels, “…the question where physics begins and mind ends or where physics ends and mind begins…” appears much less significant (almost like a pseudo-problem) than the issue of interaction (Popper et al. 1993, 171–172). The real, important issue for Popper is that minds can interact with brains (Popper et al. 1993,

---

[3] As a preliminary, note that phenomena, which are based on certain processes but cannot be reduced to (or explained by the theories of) the underlying processes are often understood as emergent.

[4] For a detailed discussion see Popper (1979).

172). This Popperian conjecture might have provoked some scholars (e.g., Lindahl and Århem 2016) to interpret the distinction between minds or consciousness (the Popperian world 2) and brains or bodies (the Popperian world 1) epistemologically, leaving the issue of whether or not consciousness is material or physical in nature an open question. The reason being, introspection suggests that consciousness is something subjective, but the activities and states of the brain are not so (Lindahl and Århem 2016, 229). Consequently, consciousness cannot be identified (in an epistemic sense) with the activities or states of the brain. This epistemic interpretation offered by Lindahl and Århem (2016) looks promising. For, though we would still have to explain—how it is possible, at least in principle, for something objective (e.g., a change in a certain neuronal electromagnetic field in the brain) to cause something subjective (e.g., an occurrence of an unpleasant sensation), and for something subjective (e.g., the occurrence of the unpleasant sensation) to cause something objective (e.g., a change in a certain neuronal electromagnetic field in the brain)—such an epistemic account wouldn't be affected by what is called the thermodynamic argument—the accusation that any action of world 2 on world 1 would violate the principle of the conservation of energy. However, what this interpretation tends to undervalue is the evolutionary significance of both world 1 and world 2. It is not evident how an epistemic reading of the Popperian distinction between world 2 and world 1 could help us approach the problem of consciousness-brain interaction from an evolutionary-biological perspective.

Conversely, Popper's hypothesis may be interpreted as providing a description of minds (world 2) being irreducibly or ontologically different from, but still interacting with, the brains (world 1). An interpretation like this is obviously prone to serious difficulties. For critics would want to know: "…how the mechanics of energy transfers work when non-physical minds move our bodies, and when non-conscious brains create conscious minds…" (Jones 2013, 11). This critical query most likely assumes that any interactions between non-physical minds and non-conscious brains would lead to violation of the first and/or second law of thermodynamics. Several attempts have been made to offer a plausible explanation of minds-brains interaction and a scrutiny of such attempts is indeed crucial for exploring any solutions to the above query. But before undertaking this task in the next section, a problem with the ontological interpretation of the Popperian distinction between minds and brains needs to be attended. The distinction between 'non-physical minds' and 'non-conscious brains' as drawn by Jones (2013, 11) seems much sharper than what Popper has proposed. Though Popper has occasionally used words such as 'immaterial' (Popper and Eccles 1977, 178) or 'incorporeal' (Popper et al. 1993, 168) to characterize minds, his emphasis on the possibility of minds' sinking into physiology is unmistakable. This is perhaps nowhere more evident than in the following remarks of Popper (Popper et al. 1993, 171–172):

> …What is very interesting is that mind may, in a sense, sink into physiology. Take a typical case, you learn to play the piano, or you learn to ride a bicycle, or you learn to drive a motorcar; in this there is a stage at which you are very conscious of everything that happens; everything is done consciously. This stage soon disappears. Mind becomes unconscious… it sinks into physiology. It sinks and becomes physiological … nobody can really deny that

> this happens, that there is a mergent process, a process where mind and brain are no longer really distinguishable.

Jones' (2013) reading of Popper reflects a common tendency among scholars (see e.g., Dennett 1979; Lowe 2000a) of ignoring Popper's non-dualistic framework while addressing the problem of consciousness-brain interactions. Popper, on the contrary, has approached the issue of interaction from an evolutionary-biological perspective that critically involves his conceptual framework of three worlds or evolutionary levels, each one having a long prehistory and none being homogeneous (Popper 1994, 111). The chief virtue of this tripartite schema presupposed in Popper's interactionist hypothesis is that it illuminates two significant points at one go. First, consciousness, as explained by Popper, not only interacts with the brain or its neurophysiological processes but also with its own objective products belonging to world 3 (Popper and Eccles 1977; Popper 1982). Close interactions among all three worlds in a mutually reinforcing way is a driving force in biological evolution, including the evolution of consciousness (Popper 1982). Second, this Popperian account of an interacting triad of evolutionary levels helps clarify several issues related to the emergence of novel, (i.e., qualitatively different) structures or phenomena.

Given Popper's distinctive emphasis on consciousness being an emergent phenomenon, and given the weakness of the above ontological and epistemological interpretations, the distinction between minds or consciousness (world 2) and brains (world 1) should better be seen as a distinction between an emergent phenomenon and its underlying, basal phenomena. The explanation of consciousness as a phenomenon emerging from, yet not fully reducible to, certain neurophysiological brain processes neither entails that consciousness is essentially something subjective nor that it is utterly non-physical in nature. The lurking doubts about how consciousness (as an emergent phenomenon) causally interacts with the brain (the underlying basal structure) would of course not immediately disappear, but in light of Popper's view of emergent evolution the aforesaid problem of interaction does appear to be more tractable. What is worth-remembering, Popper has not employed the concept of emergence merely as a productive tool for formulating significant theoretical claims about certain domains. For him, the assertion that a given phenomenon, say consciousness, is an emergent phenomenon, or that consciousness emerges from neurophysiological processes, is to say something significant, explanatory, and illuminating about consciousness and its relation to neurobiological processes.

The issue of emergence has a long history within philosophy (though its precise characterization is still contested in the existing philosophical literature), but its position within science seems both recent and tentative. The middle years of the twentieth century witnessed spectacular advances in physics and biology, particularly in the elucidation of the fundamental structure of matter (e.g. atomic, nuclear, and subatomic particle physics and quantum mechanics) and the molecular basis of biology. This progress has no doubt encouraged the scientists to believe that if the universe consists of elementary particles, and all entities are structures of such particles, then everything in the universe ought to be (in principle) explicable and predictable in terms of particle-structure and particle-interaction. This common intuition of the scientists seems to have promoted a reductionist approach—roughly speaking, the approach of explaining a phenomenon by appealing to the properties of the next level down—on the one hand, and have led to the criticism of 'emergentism' (Kim 2006a, 190)—a set of doctrines concerning the existence and characteristics of emergent properties formulated during the first half of the twentieth century—on the other.

Few would deny the power and efficacy of the reductionist approach as a methodology. As physicists have probed ever deeper into the microscopic realm of matter, the arrows of explanation will point downward (Weinberg 1994). In this way the behaviour of gases is explained by molecules, the properties of molecules are explained by atoms, which in turn are explained by nuclei and electrons. Emergentism, in contrast, though largely ignored in mainstream philosophy during the mid-twentieth century, has undergone something of a revival (as a concept) since the early 1990s (Kim 2006a, 190). However, despite the growing philosophical literature, there is little consensus on the exact content of the concept of emergence.

Let us first decide on a serviceably clear concept of emergence before examining Popper's view of consciousness as an emergent phenomenon. Kim's (2006a, 197–198) account of emergent phenomena seems clear and robust enough for the purposes on hand. Supervenience and irreducibility are considered by Kim (2006a) as two necessary conditions of emergence. Popper's explanation of emergent properties as causally dependent on but autonomous from their underlying base, though not sufficiently precise, looks quite compatible with that of Kim. While some contemporary writers (see e.g., Taylor 2015) identify an apparent tension between the features of dependence and autonomy, Popper hasn't hesitated to include these mutually non-exclusive features in his account of emergent evolution.

It is sometimes argued (e.g., Van Gulick 2001) that an emergent property does not 'supervene' on the microstructure of an object. For, an emergent property of a whole is not determined by the properties and relations characterizing its parts. If the connection between an emergent mental phenomenon, say pain, and a certain configuration of neural conditions from which it emerges, is so irregular or coincidental, one may ask following Kim (2006a), what reason could there be for arguing that pain 'emerges' from that neural conditions rather than another? Emphasizing supervenience as a necessary component of emergence, Kim (2006a, 193) states the condition of supervenience as follows: If property M emerges from properties $N_1,\ldots, N_n$, then M supervenes on $N_1,\ldots, N_n$.[5]

Supervenience, though necessary, is not sufficient for emergence. The surface area of a sphere supervenes on its volume, but it does not emerge from it. On the contrary, according to most advocates of emergence (including Popper), consciousness both supervenes on and emerges from physical-biological conditions. Thus, something must be added to supervenience to yield emergence. The basic idea explained by

---

[5] Kim (2006a, 193) defines supervenience as follows: to say that M supervenes on $N_1, \ldots, N_n$ is to say that any system that has the base properties $N_1, \ldots, N_n$ will necessarily have the supervenient property M.

Kim (2006a, 194) is that if M emerges from N_1, …, $N_{2,}$ then although M supervenes on the N_s, M is not reducible to, explainable in terms of, predictable on the basis of, or derivable from, the N_s. Therefore, according to Kim (2006a, 197), property M is emergent from properties N_1, …, N_n only if (i) M supervenes on N_1, …, N_n, and (ii) M is not functionally reducible with N_1, …, N_n as its realizers. Thus, supervenience and irreducibility in the sense explained above are two necessary conditions of emergence. Given these necessary conditions, it is immediately clear that emergent properties must have some causal power, and this includes their capacity for projecting causal influence downward, affecting the course of events at a purely physio-chemical level.

This very idea of an emergent structure operating causally upon its sub-structure in a direct downward fashion appears incoherent to many, including Kim (2006b). We seem to understand upward causation well, that is, how the sub-structure of a system cooperates to affect the whole system. Difficult is to envisage how a higher-level, emergent structure operates causally upon its sub-structure. Resisting the extreme reductionist tendency to explain everything in terms of causally interacting elementary particles, Campbell (1974) showed how a change or action from above can also affect the set of sub-structures. For example, the average velocity of a group of atoms can influence the average velocity of the neighbouring groups of atoms and can thereby influence the velocities of many individual atoms in the group. Taking his cue from Campbell, Popper (1978, 348) has tried to explain downward causation as ‘selection’ operating on the randomly fluctuating elementary particles. The randomness of the movements of the elementary particles provides the opening for a higher-level structure to interfere from above (Popper 1978).

The most recurrent and profound problem relating to emergent properties’ capacity for projecting causal influence downward arises from the closed character of the physical domain having the following implication: if a physical event has a cause, it has a physical cause; and if a physical event has an explanation, it has a physical explanation (Kim 2006a, 199). Arguably, this causal closure principle of the physical domain is presupposed by most working scientists, including of course physicists. If the scientists encounter a physical event for which they are not able to identify a physical cause or explanation, it is highly unlikely that they will consider positing a non-physical cause to explain it. To deny this principle basically amounts to denying the in-principle completability of theoretical physics (Kim 2006a, 199–200).

Popper (Popper and Eccles 1977, 15) has never questioned the physicalist premise that nothing can happen unless permitted by the physical laws and by the preceding (physical) state. He has only objected to what we commonly infer from it. From the point of view of human knowledge, he (Popper and Eccles 1977, 15) has cautioned us, it would be misleading to conclude from the above seemingly indisputable physicalist premise that the future is and always was foreseeable (in principle, at least). To combat the physicalist view of the completeness of the physical world Popper (Popper 1979a) has come up with a simple argument based on his conjecture of three worlds. He was quite sure that the physicalists (or materialist monists) would not readily accept his pluralist conjecture (Popper 1979a). They would assert either there is only world 1 or even if there is a world 2 or a world 3, neither can act on world 1 (Popper and Eccles 1977, 51).

Nevertheless, Popper's argument begins by stating what seems undeniable, namely, that we live in a physical world (world 1) which has been greatly changed by making use of science, i.e., scientific conjectures or theories (world 3 entities) as instruments of change (Popper 1979a). Next, assuming 'kickability' as a sufficient condition for reality (Popper 1994, 47), the argument concludes by asserting that scientific conjectures or theories can, therefore, exert a demonstrably causal or an instrumental effect upon physical things (Popper 1979a, 8–9). The very existence of objective problems together with the fact that its discovery and solution by means of scientific conjectures or theories may lead to obvious changes in world 1 implies that the physical world 1 is neither closed nor complete but open towards world 3 with world 2 acting as an intermediary (Popper 1979a).

The basic outcome of Popper's conjecture of three interacting worlds is no different from 'emergentist pluralism' (Ellis 2006, 85)—a philosophical position that assumes the emergence of many levels of reality in the natural world. In addition, the objects at these various levels that can be shown to have a causal effect in the material world of particles are assumed to have their own types of reality. These include human concepts, plans, intentions, information, emotions, as well as socially constructed features such as chess rules (Ellis 2006, 84–85). Ellis' (2006) analysis of emergence is almost a replica of that of Popper's, excepting its key philosophical implication. Ellis (2006, 84–85) seems unsure about whether true emergence is ever possible. That is, whether the creation through physical and biological processes of completely new types of structure without any kind of precursor is the creation of a completely new kind of order, or whether emergence in the physical world is simply the realization of pre-existing potential and hence not a truly creative event. Ellis' (2006) worry reminds us of the common argument of the critics of emergence, namely, that evolution is a fact, but evolution cannot be 'emergent' or 'creative.'

Popper, in contrast, has described emergent phenomena as 'altogether unpredictable' (Popper and Eccles 1977, 22) and compared their novelty with that of a great work of art (Popper and Eccles 1977, 22). Two promising insights regarding the possible emergence of genuinely novel structures may be distilled from Popper's writings. The first one relates to his (Popper and Eccles 1977, 25) firm belief that there can be invariant physical laws and emergence of new properties as the former is not sufficiently complete and restrictive to prevent the latter. Even if we admit that newly emergent entities create new fields of propensities, one of Popper's critics asks, can we really escape the idea of preformation or several preformationist possibilities (Popper and Eccles 1977, 31)? Popper's answer to this question is simple but noteworthy. That we may have an infinity of such open preformationist possibilities is reason enough to dispense with preformationism (Popper and Eccles 1977, 31).

The second Popperian insight relates to his recognition of a 'whole' as distinct from 'mere heaps' (Popper and Eccles 1977, 20). A 'whole,' for instance, a living organism, is an emergent macro-structure which is more than a mere heap or sum of its parts in the sense that though it emerges out of interactions between its underlying, constituting parts is neither completely predictable nor reductively explainable (at least in any straightforward manner) in terms of those parts (Popper and Eccles 1977, 21). The examples of living organisms or other such emergent structures, as per Popper, make the existence of downward causation obvious, and by implication, challenge the 'complete success' (Popper and Eccles 1977, 20) of any reductionist program.

The tension between reductionism and emergence is easily visible. For any attempt to minimise this tension, a brief overview of weak and strong versions of both emergence and reductionism may be helpful (Davies 2006b, xi). It goes without saying that the physicists' ability to break apart atomic particles into smaller and smaller fragments and to probe ever deeper into the microscopic realm of matter is essential for our understanding of the properties of matter or the fundamental forces that shape it. One might be tempted to know, whether the reductionist account of nature is merely a fruitful methodology—a weak form of reductionism known as methodological reductionism—or whether the whole really is, in the final analysis, nothing but the mere sum of the parts. This later, stronger form of reductionism is sometimes known as ontological reductionism (Davies 2006b, xii). While many scientists, particularly physicists, are self-confessed strong, ontological reductionists, a minority of scientists (see e.g., Davies 2006a; Ellis 2006) today find the basic claim of ontological reductionism contestable.

The contrast between 'weak' and 'strong' versions of emergence is also to be noted. The position of weak emergence recognizes that one may not be able to deduce merely from the principles that govern a class of systems how a specific individual system will in fact behave (Davies 2006b, xii). However, direct inspection or simulation may enable us to determine the behaviour of complex systems, such as human behaviour or even that of a simple organism (e.g., a bacterium). Strong emergence, on the other hand, is a far more contentious position as it asserts that the micro-level principles are quite simply inadequate to account for the system's behaviour (as a whole). Evidently, strong emergence cannot succeed in systems that are causally closed at the microscopic level, because there is no room for additional principles to operate that are not already implicit in the lower-level rules. For instance, a closed system of Newtonian particles cannot exhibit strongly emergent properties, as everything that can be said about the system is already contained in the micro-level dynamics including the initial conditions (Davies 2006a).

There are 'three loopholes' that make strong emergence conceivable (Davies 2006b, xii). The first is if the universe is taken as an open system, the system as a whole would then be determined partially by the micro-level dynamics and partially by the constraints imposed by the external, global principles—principles which may 'soak up' the causal slack left by the openness (Davies 2006b, xii). The second possibility comes into sight if the system is interpreted as non-deterministic—quantum mechanics being the popular example—and unique rather than belonging to a homogeneous ensemble. The final possibility arises if the laws of physics operating at the base level are understood as intrinsically imprecise due to the finite computational resources of the universe. Similar possibilities—commonly perceived as unorthodox departures from standard physical theory—have also been considered by Popper (Popper 1974, 1982, 1988) long ago. The very fact that such possibilities are being assessed by contemporary scientists increases the prospect for a stronger argument for emergent evolution.

Nevertheless, the important challenges that no ardent supporter of emergence (not to mention Popper) can avoid are as follows: an emergentist must either provide sufficient and compelling reasons for rejecting the causal closure principle or else show that downward causal efficacy of irreducible emergent properties is consistent with physical causal closure (Kim 2006a, 199–201). If any emergentist accepts the challenge and gives up the causal closure principle, then a further challenge would be to offer a credible explanation (that goes beyond supervenience and irreducibility) of how minds are related to the activities of the brains. Saying that minds emerge from brain-processes but are not reducible to them doesn't say enough about their relationship. Whichever option is adopted, the 'friends of emergence' (Kim 2006a, 201) are indeed in trouble because of the philosophers' lingering, unargued, uncritical allegiance to the causal closure principle.[6]

Just as the friends of emergence, as per Kim (2006a), have only two choices, the majority of the physicists today are also faced with two challenging options: either to extend the scope of physical description to encapsulate higher-level causal effects, such as the causal effects of conscious plans and intentions, emotions or thoughts in the real physical world; or to decide that these kinds of issues are outside the province of physics, which properly deals only with inanimate objects and their interactions (Ellis 2006). Whichever option is adopted, their ambitious aim of providing a causally complete description of all interactions that affect the real physical world or a 'theory of everything'—a unified theory of fundamental forces and interactions such as String Theory (see, e.g. Greene 1999)—seems to be in trouble. At minimum, physics must be related somehow to the world of thoughts and feelings before it can make any claim to provide causal completeness—which presumably a true 'theory of everything' aims at.

A review of the widely accepted causal closure principle seems necessary for any discussion about the challenges faced either by the friends of emergence or by the critics of emergence. As things now stand, the task before Popper (as an emergentist) is to explain how minds or consciousness as novel, emergent phenomena having distinct, irreducible causal powers arise from and interact with the brain without any violation of the causal closure principle of the physical domain, or more specifically, of the laws of classical physics (e.g., first and/or second law of thermodynamics). This very question—how consciousness and brain can possibly interact—remains one of the most 'mysterious' and 'intractable' problems in philosophical investigations of the mind-brain relationship (Libet 1994, 120). The following section attempts to show how far the Popperian hypotheses can reasonably demystify the problem of consciousness-brain interaction.

---

[6] For a review of different formulations of the causal closure principle see (Lowe 2000a).

3 Does Consciousness-Brain Interaction Necessarily Violate Physical Laws?

In his 1977 book *The Self and Its Brain,* Popper has only briefly addressed the problem of detecting neural activities related to consciousness (Popper and Eccles 1977, 117–120). Hypothesizing about the possibility of there being a one-to-one relationship between certain conscious experiences and certain brain processes Popper has commented that to be linked with consciousness—'a teaming process of unimaginable complexity' (Popper and Eccles 1977, 120)—the whole brain must be in high activity. His later writings suggest that both in the phylogeny and in the ontogeny of humans, self-consciousness appears with the higher functions of language and interaction occurs between the self and the speech centre of the brain (Popper 1994, 131–132). His latest mental force field hypothesis introduces a striking analogy between minds and physical forces and indicates a new possibility of how minds and brains can interact (Popper et al. 1993).

In a wide-ranging interview (Popper et al. 1993) published near the end of his life, Popper has come up with the proposal that conscious minds should be understood literally as a field of forces. To somewhat demystify the existence of minds he has drawn attention to several similarities between conscious minds and physical forces (or fields of forces). Minds, like forces, have at least six properties: (i) they are located in space and time (ii) they are unextended in space but extended in time (iii) they are incorporeal but existing only in the presence of bodies (iv) they are capable of acting on bodies (v) they are dependent upon bodies, and (vi) they are capable of being influenced by bodies' (Popper et al. 1993, 168). Later he has added two more properties that minds have in common with physical forces, namely (vii) minds are intensities and, they have (viii) minds have extension through a span of time (Popper et al. 1993, 168).

Objecting to our common understanding of forces as 'mere appendices to matter' (Popper et al. 1993, 169) Popper has argued that the forces, though related to biochemical substances or physiological entities, can, apparently, obtain a certain autonomy and independence from these sheer substantial processes with which they are related and with which they interact. Similarly, minds—something like a complex of forces—can make themselves independent of the physiology and can have a certain amount of life on their own (though the physiology is always present). This (partially) explains why minds or conscious processes (world 2) are seen by Popper as emerging from, but not fully reducible to their physical-chemical basal structure (world 1).

Popper has also developed an idea of mind as (at least partly) a force field. The complicated electro-magnetic wave fields which are part of the physiology of our brains, represent the unconscious parts of our minds, and conscious minds—our conscious mental intensities, our conscious experiences—are capable of interacting with these unconscious physical force fields, especially when the problems to be solved require what we call 'attention' (Popper et al. 1993, 179). Our force-like minds (or mental experiences) always point to something, always intend to bring something about. This characteristic feature has been referred to by Popper as ‘intentionality’ (Popper et al. 1993, 172), which like a vector always points to something and has the power, to bring something about.

This mental force field hypothesis of Popper seems quite a thought-provoking attempt to account for the causally effective nature of minds or consciousness. Conscious processes—assumed to emerge as a function of appropriate neural activities in the brain—seem capable of acting back on certain neural activities through unconscious (i.e., purely physiological) physical force fields (Popper et al. 1993, 172). The argument implicit in this latest hypothesis of Popper is as follows: if it is acceptable that physical forces can influence bodies, and bodies can influence physical forces, then, the interactions between consciousness (which is quite similar to physical forces) and the neuro-physiological activities of the brains would not be too mysterious or difficult to explain (Popper et al. 1993).

During the last twenty-five years several elaborate field theories of mind have been proposed (for a review, see Jones 2013). In many of these theories, electromagnetic fields of the central nervous system are taken to be crucial to the explanation of conscious experiences. Different components of the brain’s electromagnetic field are understood to be relevant to consciousness, and in fundamentally different ways—e.g. as being identical with or being the substrate of consciousness. The currently competing theories of the neural basis of human consciousness vary considerably as to which brain areas and activities are suggested to cause conscious experience. Popper’s mental force field hypothesis no doubt supplements these theories by emphasizing important similarities between minds and physical forces and indicating some possible ways of minds-brains interactions.

In spite of these interesting suggestions of Popper, the following question might still arise: is there any compelling reasons for believing that consciousness does causally interact with certain neural processes of the brain (through unconscious force fields) rather than being merely supervenient on those brain processes? Arguably, the strongest objection against any form of interaction concerns its explanation of the causal efficacy of consciousness in a way that entails violation of physical laws (e.g., the first and/or second law of law of thermodynamics). The said objection rests on the persisting belief that consciousness can act on the brain only if the physical realm is causally open. Popper’s (1984a, 21–22) own reply to the problem mentioned above, namely, does the very possibility of consciousness-brain interaction necessarily violate physical laws, is quite promising, but Averill and Keating’s (1981) analysis of the same is more instructive for the present purposes.

In the history of physics both the first and the second law of thermodynamics have appeared in different forms. Rudolf Clausius, for instance, formulated the first law as stating that ‘in any closed system (a steam engine, for example) the total amount of energy is constant’ and the second law as stating that ‘heat cannot pass from a colder to a hotter body on its own accord; for this to happen some external cause must come into operation’ (Ronan 1983, 447). The point we need to note is, laws of (classical) physics are open to different interpretations.

In *The Open Universe*, Popper has referred to the first law as ‘the law of conservation of energy’ and the second as ‘the law that asserts that entropy can only increase’

(Popper 1988). He (Popper 1988) has argued that the first law holds only with (more or less) good approximation for organisms, since living organisms are not closed systems (like steam engines). He has also hypothesized about the existence of 'purely mental forms of energy, convertible into electrochemical forms' (Popper 1984a, 21). What's more, the possibility of non-energetic influences upon energetic processes has been considered by Popper on account of certain interpretations of de Broglie's particle-wave theory postulating the existence of empty pilot waves that can interfere with non-empty (energy-carrying) waves (Popper 1984a, 21–22). The second law, according to Popper (1984a, 22) has been refuted by Brownian movement.

Averill and Keating's (1981) explanation of the problem relies upon a standard textbook of mechanics, namely, Herbert Goldstein's *Classical Mechanics,* to deal with theoretical physics. Interestingly, Averill and Keating's (1981) analysis avoids making any reference to closed systems, but it does make several references to the presence or absence of external, especially mental, force. Considering Popper's late construal of the force-like nature of minds, Averill and Keating's (1981) discussion about mental force (and not mental energy) seems most suitable for our purposes. The basic claim of the interactionist position has been stated by Averill and Keating (1981, 102) in the following way: consciousness (or a non-physical mind) can initiate behavior by exerting a force which moves a brain-particle.

In the opinion of some critics (e.g., Dennett 1978, 252; Corman 1978, 274) of interactionism, first of all, the motion of the brain-particle resulting from such a non-physical force would change the total energy and/or the total linear momentum of the brain; and secondly, if the resulting motion changed the total energy and/or the total linear-momentum of the brain, then either the conservation of energy law or the conservation of linear-momentum law would be violated. In addition, the assertions of Cornman (1978, 274) mentioned below are also note-worthy:

(1) If the mind exerts a force F on the brain, and F changes the resistance at certain synapses, then the total linear-momentum of the brain is changed due to F.

(2) If the total linear-momentum of the brain is changed, then 'some net external physical force' affects the brain.

Let us now consider the text-book formulation of the law of conservation of linear-momentum for a system of particles. In Goldstein's *Classical Mechanics* the law has been stated as: 'If the total external force is zero, the total linear-momentum is conserved' (Averill and Keating 1981, 103). Clearly, Goldstein's interpretation of the law does not entail (2). For, it has no such implication that if the total linear-momentum of a physical system, say, the brain, is changed, then 'some net external physical force' is affecting the brain (Averill and Keating 1981, 103). Besides, notice that the said law as formulated by Goldstein may hold for all kinds of forces, regardless of their source, and not just the 'physical forces' whose source is a physical object.

The possibility of the (non-physical) mind's exerting a force F on the brain is thus not denied by the law of conservation of linear-momentum as articulated in Goldstein's classic text. An interactionist may retain the law and reject (2). Cornman's (1978, 274) has rightly inferred that any force exerted by the non-physical mind on the brain has to be external to every physical system (including the brain) and therefore such a force would not be one of the 'only appropriate physical forces,' such as gravitational forces (which require mass) and electromagnetic forces (which require electrical charge). However, it does not necessarily follow from Cornman's observation that such a force is 'not appropriate to mind-body interaction' (Averill and Keating 1981, 103).

Cornman's basic error, according to Averill and Keating (1981, 104), consists in using a much stronger statement of a physical law than is necessary for the development of physics. Moreover, this stronger formulation of the said law has (metaphysical) implications which beg the question against interactionism. In contrast, Goldstein's interpretation of the same law does not require that the source of a force which changes the total linear-momentum of a system must be physical, and hence does not have any implication like the following:

(3) If X exerts a force F on a physical system S, and the total linear—momentum of S is changed due to F, then X is physical.

An interactionist can, therefore, reject (3) but retain the conservation law for linear—momentum (Averill and Keating 1981, 105).

Consider another example of a physical law. The dilemma often faced by many interactionists is whether to reject the principle of conservation of energy, or to show how energy may be conserved in the brain when consciousness exerts a force on the brain. The popular interpretations of law of conservation of energy—also known as the First Law of Thermodynamics—has the following implication (Averill and Keating 1981, 106):

(4) If X exerts a force F on a physical system S, and the total energy of S is changed due to F, then X is physical.

In the textbooks of physics, the said law is formulated as (Averill and Keating 1981, 106): $\Delta U + L = Q$.

where ΔU is the change in energy of a system, L is the work done by the system during the change ($-L$ is the work done on the system), and Q is the heat flowing into the system during the change. More generally, Q is the amount of energy that is received by the system in forms other than work (Averill and Keating 1981, 106).

This textbook-formulation of the law does not entail (4) as no clue is provided about the source of the force that does the work ($-L$) on the system. Hence it neither assumes that there is a change in energy in the source of the force, nor that the source of the force is part of a physical system. Here again the same kind of error is repeated—the use of a statement of a physical law that has question-begging implications and is stricter than is required for scientific research (Averill and Keating 1981, 106). An interactionist faces no difficulty in rejecting (4) and retaining the law as formulated in one of the standard textbooks of mechanics.

Now, if, as explained by Averill and Keating (1981), the possibility of the non-physical mind's acting on the brain entails no violation of the physical laws mentioned above, then the main objection against interactionism, which now seems to rest on the philosophers' preconceived beliefs about physical laws, loses ground. Given that physics doesn't necessarily rule out the key claim of interactionism, one may want to ask the critics (of interactionism) whether or not it is a prejudice to reject the very conceivability of an interactionist hypothesis without examining the conventional formulations of physical laws that the hypothesis arguably conflicts with.

The mistake plaguing the conventional interpretations of the said physical laws is of importance no doubt, but that doesn't stop one from inquiring if Popper's (non-dualist) interactionist hypothesis offers anything new. Are there reasons different from those already discussed in the existing literature (e.g., Lindahl 1997) for revisiting his decades-old hypothesis? The originality of Popper seems most evident in his attempts at explaining the evolution of consciousness' (adaptive) functions from lower to higher organisms. Popper is not the first one to formulate an evolutionary (philosophical) argument for mind-brain interaction (see e.g., James 1879), but he is one of those rare philosophers who has reflected on the evolutionary origins of the mind, both pre-human and human, has identified the initial appearance of mind-like behaviour (e.g., alertness, eager) of organisms very early in evolution and hypothesized about how the mind-like behaviour of organisms gradually evolved into exploratory (trail-and-error), partly-conscious behaviour (e.g., anticipation of future needs, preference for certain kinds of food or locations for breeding, actively searching for new ecological niches) of higher organisms and how that partly-conscious behaviour of higher organisms developed into conscious behaviour (e.g., creation of world 3 entities) of human beings, though we cannot determine exactly when (Popper 1982). What's more, Popper (1982, 45) is probably the first philosopher to argue that the emerging minds—pre-human and human—play an active part in biological evolution and especially in their own evolution.

Popper has developed his argument by presenting a neo-Darwinian interpretation of the process of adaptation. Adaptation, according to him (Popper 1982) is a process based upon reciprocity and the activities of the living organisms. Activity is a movement with an aim and, therefore, without aims, such as striving for food, or for a higher or lower temperature, adaptation is inconceivable. The living organisms strive for food or for a higher or lower temperature by actively selecting and changing their own environment, their ecological niche, like birds build nests, or humans construct structures that broadly enhance their evolutionary fitness. What is usually perceived as the (more or less) passive reactions of the organisms to the environmental stimuli has been explained by Popper (1984b, 244) as exploratory actions of the organisms. Through many trials over millions of generations the organism learns to exploit environmental change as a stimulus and invents the ability to react to it as stimulus (Popper 1984b, 244). Simply put, what turns something into a stimulus is the 'eagerness' of the organism to respond or react according to its internal state (Popper 1984b, 244–245). Popper (1982, 40) has described such active, exploratory behaviour as 'mindlike' or 'partly conscious' and argued that with more and more complex forms of life conscious aims (e.g., preference for certain locations for breeding, or for certain types of food or for certain kinds of mates) appeared from these mind-like, exploratory behaviour.

Promising though it may seem, Popper's non-dualist interactionist hypothesis involves some speculative premises, which are neither conclusively verifiable nor sufficiently precise. His speculation about all[7] living organisms being active explorers or about how the mind-like, exploratory behaviour of lower-organisms effectively evolves into conscious behaviour of higher organisms is a case in point. The virtues of Popper's tentative, old-fashioned (non-dualist) interactionist hypothesis, however, lie in its openness to critical experimental-archaeological scrutiny (if not to strictly scientific tests). The next section discusses in what way Popper's hypothesis can be subjected to experimental-archaeological investigations.

4 The Archaeological Implications of Popper's (Non-dualist) Interactionist Hypothesis

The epistemic underpinnings of Popper's (non-dualist) interactionist hypothesis is rooted in its 'anti-behaviourist' and 'anti-psychologistic' character (Popper 1979b, 114). The said hypothesis presupposes the crucial distinction between knowledge in the objective sense (world 3) and knowledge in the subjective sense (world 2) on the one hand, and indicates how the problems concerned with the acts of production differ from the problems concerned with the (objective) structures or products (Popper 1979b, 114), on the other. The problems connected with the products or structures (world 3) themselves, according to Popper (1979b, 114) are more fundamental because they illuminate production behaviour and psychology (world 2). The conventional approach of studying minds or consciousness (world 2) for acquiring information about their (objective) products has a scientific appeal because it proceeds from causes to effects. Popper (1979b, 112), in sharp contrast, has argued that the reverse approach is more significant. That is, one can learn more about minds and their behaviour (world 2) by examining the effects generated or products caused by minds. This observation of Popper—that the objective approach of examining world 3 can help throw light upon world 2 mental or conscious processes—holds immense significance for archaeological investigations.

The voluminous record of stone-tools (e.g., a hammer-stone or a hand axe) shaped and used by the prehistoric hominins is widely seen by archaeologists today as the most enduring source of evidence for the initial emergence of (some form of) hominin conscious-cognitive behaviour (see, e.g., Wynn and Coolidge 2016; Moore and Perston 2016). There is a considerable archaeological literature on the cognitive dimensions of specific hominin technical activities such as stone knapping (e.g., De Beaune 2004; Nowell and Davidson 2010) or, more generally, stone-tool making (e.g., Stout and Chaminade 2009; De la Torre 2011; Wynn 2009; Wynn and Coolidge 2016, 2017). These studies strengthen the archaeological intuition that the stone-tools shaped and used by ancient hominins played a seminal role in the evolution of the early hominin conscious-cognitive abilities. However, within archaeology and the

---

[7] There are a few cases such as the adaptation of bacteria to penicillin where the catastrophic changes do not allow the organism to be active as all members of the population are killed (Popper 1982, 41).

study of hominin evolution, stone tools are typically described as mere end-products or by-products of the hominin minds (or conscious-cognitive abilities). Evidently, the causal arrow assumed in this standard perception is one way—from minds (or conscious-cognitive abilities or processes) to tools or cultural products. Among major issues that have arisen in the past few decades are questions regarding the critical role of stone-tools in the evolution of hominin mental or conscious-cognitive faculties.

Experimental-archaeological investigations (e.g., Stout et al. 2008; Stout and Chaminade 2007, 2009, 2012) into the neural corelates of prehistoric (lower Palaeolithic) tool-behaviour leave the impression that conscious-cognitive processes are correlated, if not totally identical, with certain neurophysiological processes of the brain. This experiment-based approach uses new experimental techniques to unravel the connection between stone-tool production and brain-processes. On the other hand, some present-day researchers known as cognitive archaeologists (e.g., Malafouris 2013) draw inspiration from the philosophical hypothesis of the extended mind (Clark and Chalmers 1998) and have interpret cognitive-processes not as essentially brain-bound but as processes transcending the cranial boundary and incorporating extra-somatic, environmental resources. This latter (more theoretical) approach puts special emphasis on the 'explanatory and transformative power' (Malafouris 2013, 57) of such extra-somatic resources (e.g., stone-tools) and paves the way for a deeper interaction between contemporary philosophical and archaeological research on the evolution of consciousness.

Popper's (non-dualist) interactionist framework let the archaeologists explain and emphasize the efficacy and transformative potential of human or hominin products without adhering to the philosophical thesis of the extended mind, objections to which are neither rare nor negligible (see e.g., Rupert 2004). It isn't archaeologically essential to interpret stone-tools as genuine extensions of hominin conscious-cognitive processes for emphasizing the critical impact of the former on the latter. This insightful tripartite account of Popper, crucial for understanding the prehistory of minds, is completely ignored by contemporary archaeologists (e.g., Malafouris 2013). Besides, Popper's description of consciousness as an emergent (biological) phenomenon—dependent upon yet not completely reducible to underlying neurophysiological processes—also enables the archaeologists to avoid the controversies related to 'neurocentrism' (Malafouris 2009, 258)—a physicalist-style attempt to reduce all the properties of cultural products to the properties of the brain—without falling into the old Cartesian trap of thinking about the brain as something physical and consciousness as purely non-physical.

Following Eccles (Popper and Eccles 1977, 450) one might want to ask Popper: how far back in the human prehistory can we recognize the origin, the most primitive World 3 entities? The very beginning of World 3, in the view of Popper (Popper and Eccles 1977, 451), is to be detected not in the earliest tools made by the hominins but in the initial development of some form of hominin linguistic behaviour. For, prior to some (primitive) kind of linguistic behaviour, stone-tools could not be regarded as objects of 'criticism and of deliberate improvement' (Popper and Eccles 1977, 451).

Experimental-archaeological reconstructions (see, e.g., Toth 1987; Toth and Schick 2009; Stout 2011) of the manufacturing-process of prehistoric Oldowan stone-tools reveals that Oldowan raw materials had been examined at the source, selected stone resources had been transported for initial flaking at a second location, and selected flaking-products had been transported for use at a third location. Wynn and colleagues' (Wynn et al. 2011, 195) more recent research on the long-distance transport of various raw materials is also indicative of the Oldowan hominins' capacities for high-level planning or anticipatory behaviour and of these ancient hominins' selectiveness in the use of those raw materials. Capitalizing on the archaeological evidence of hominin preference for certain types of stones as suitable for flaking one might take the deliberate hominin rejection of certain kind of raw materials as some sort of criticism or as a 'forerunner of criticism' (Popper and Eccles 1977, 451). Conversely, if criticism proper is supposed to arise only with some kind of linguistic behaviour, and if language-processing is assumed to be a functionally specialized and anatomically discrete module within the brain, then Popper's speculation about hominin linguistic competence being critical for hominin tool behaviour and consequently for the origin of world 3 definitely calls for additional supporting evidence.

Fortunately, experimental-archaeological data informing possible evolutionary connections between hominin linguistic and tool behaviour are available today. Three possible types of co-evolutionary interaction involving shared neural substrates, shared social context and shared reliance on general capacities are highlighted in the archaeological literature (e.g., Stout 2010). The intersection of language and manual-praxis networks in Broca's area—originally identified as specifically responsible for the faculty of spoken language—now provides one of the best-known examples of complex functional-anatomical overlap in human neocortex (Stout and Chaminade 2012). It is also recognized that the frontal 'language-relevant' cortex—extending across the entire inferior frontal gyrus (IFG)—contributes not simply to linguistic functions (e.g., the comprehension and production of syntactic, semantic or phonetic structure), but to a range of non-linguistic behaviours from object-manipulation to sequence-prediction (Stout and Chaminade 2012). Such evidence of functional-anatomical link between hominin tool-behaviour and language-competence, indicating (though not proving) specific co-evolutionary relationships between them, appears broadly compatible with the Popperian assumption that an objective, 'criticizable' world 3 (Popper and Eccles 1977, 451) probably co-evolved with development of hominin linguistic behaviour.

To further dissect the tripartite interactions of the distinct evolutionary levels one may consult recent research on neuroplasticity and the plasticity of human minds. The plasticity exhibited by human cortical maps—not only during the early-developmental period when synaptic densities are maximum in most brain-regions but also during adulthood—is associated with anatomical and not, as traditionally assumed, only with functional changes. Stout and Chaminade's (2007, 2009) experiments using positron emission tomography (PET) are worth mentioning here. Their (Stout and Chaminade 2007, 2009) comparative assessments of previously inexperienced subjects making Oldowan-style stone tools both before and after completing four-weekly practice sessions in stone-tool manufacture demonstrate noticeable functional changes in brain activation patterns following the practice sessions. The 2006 study of Hihara and colleagues also shows how two weeks of tool-making training forges a novel cortico-cortical connection linking the intraparietal area and temporoparietal junction (TPJ). These results attest to the possibility of interactions between neuro-physiological (i.e., world 1) processes and stone tools (world 3 products) via training or learning (a world 2 conscious-cognitive process).

Mithen and Parsons (2008) have argued, in a similar vein, that the human brain is being continuously re-shaped, re-wired, and re-modelled under the influence of cultural practices. Their analyses of the brain-anatomy of skilled musicians and non-musicians show how prolonged instrumental practice leads to an enlargement of the hand-area in the motor cortex. Such evidence of neuroplasticity is an important driving force behind their understanding of the human brain as a dynamic bio-cultural system or 'an artefact of culture' (Mithen and Parsons 2008, 415) subject to constant functional and structural changes. The possibility of three-way interactions among brain-processes (world 1), musical instruments (world 3 products) and instrumental-practice or training (a world 2 conscious-cognitive process) once again seems implicit in this novel construal of the human brain.

True, the evidence of neuroplasticity is not sufficient for establishing (causal) interactions among three Popperian worlds; but it does hint at the causal openness of the (world 1) brain-processes. If, as studies (e.g., Mithen and Parsons 2008; Stout and Chaminade 2007, 2009) suggest, the volume of the corpus callosum of a pianist seems connected with prolonged hours of piano-practice, or functional changes are visible in brain activation patterns following four-weekly practice sessions in stone-tool making, then the evolutionary impact of several long-term cultural practices in the structural and functional modifications of human (or hominin) brains would certainly be undeniable. This could be one of the reasons why the developing field of neuro-archaeology seems committed to an interactionist (and not a reductionist) view of mind-brain relations (Malafouris 2009, 258).

Jeffares' (2010) provocative suggestion that the ancient stone tools—often seen as mere end-products of extinct hominin minds—can, in some cases, causally trigger cognitive processes, or new cognitive capacities also hints at the plasticity of human minds. His (Jeffares 2010) vision of the critical role of the first recognisable stone tools in structuring early hominin cognitive processes has been shared by other contemporary archaeologists. For example, Malafouris' (2015), distinctive emphasis on the extraordinary projective plasticity of the mind and its reciprocal openness to cultural influence (and variation) attracts notice. Being a product of evolution, the human mind is undoubtedly constrained by several inherited genetic structures, brain circuits etc. but these genetic constraints cannot determine its developmental trajectory a priori (Malafouris 2010). Popper's conjecture about the causal openness of world 2 coheres with the (cognitive) archaeologists' hypothesis of the remarkable plasticity of the mind.

There exists credible, though not conclusive, experimental-archaeological evidence today for contending that (i) the (human) brain is not a long-evolved, fixed biological entity but an evolving plastic organ actively interacting with the physical environment (instead of passively adapting to it), with the conscious-cognitive processes (emerging from it) and with the extra-somatic products brought out by those processes; and that (ii) the human mind (despite its neuronal and somatic bases) shows remarkable plasticity in relation to the wider ecological, social, and technological environment. In light of this recent experimental-archaeological evidence, Popper's thoughts on the complex, tripartite, interactions among world 3 (e.g., stone-tools), world 2 (e.g., conscious-cognitive processes) and world 1 (e.g., the brain-regions associated with tool making activities and learning) do not seem mere guesswork.

An important clue about a close connection between prehistoric tool-behaviour and the development of hominin conscious-cognitive traits may also be found in a relatively old hypothesis of Washburn (1978). Washburn (1978, 201) has argued that technological progression from no stone-tools to simple Oldowan tools to skilfully shaped and increasingly refined Acheulean bifacial cutting-tools is correlated with the doubling or, as Stout (2011) more recently suggests, nearly tripling of hominin brain size (i.e., endocranial volume). Assuming Washburn's (1978) hypothesis to be correct, one may also expect to see an increase in the neurological complexity of hominin brains. Since fossil record of direct evidence for evolutionary changes in gross neural anatomy remains scant, Washburn (1978, 202) has concluded by simply suggesting that increasing brain-size does seem to be correlated with the increasing complexity of stone-tools over hundreds of thousands of years. Today, Orban and Caruana's (2014) recently found functional magnetic resonance imaging (fMRI) data, which correlate the presence of a new neural apparatus located in left anterior supramarginal gyrus (aSMG)—a region of the brain most likely involved in the execution of tool actions—with the emergence of *Homo habilis* (arguably the principal Oldowan tool maker), seem to lend support to Washburn's (1978) hypothesis.

The relationship between increased brain-size and lower Palaeolithic technological change, however, is a matter of ongoing debate and it might not be as direct as Washburn (1978) has supposed. Tennie and colleagues' (2017) study of the tool-making abilities of the relatively small-brained *Homo floresiensis* is a case in point. But the mere facts about the above case cannot completely overthrow Washburn's (1978) conjecture of a roughly parallel occurrence between the two most striking trends in hominin evolution, namely, the growing sophistication of stone-tools over hundreds of thousands of years and the nearly three-fold increase in hominin brain size accompanied by a possible upsurge of neural resources.[8] Considering the dominant view of the brain as the primary source of conscious-cognitive capacities, and drawing upon the experimental-archaeological evidence of increasing brain-size as well as neural complexity, some connection between the gradual development of prehistoric flaked-stone tools and the evolution of early hominin conscious-cognitive faculties becomes conceivable.

What is more, the possible (evolutionary) connection between hominin tool-production and hominin conscious-cognitive abilities may plausibly be explained

---

[8] Increasing neural complexity indicates not an increase in the absolute numbers of neurons but novel neural connections among existing neurons.

using Popper's (non-dualist) interactionist hypothesis. Early hominin conscious-cognitive traits might be interpreted, following Popper, as not merely emerging from specific hominin brain-features (most likely involved in stone tool-related activities) but also causally acting on these generating sources.

Most of the grounding philosophical assumptions on the evolution of minds or consciousness (world 2) are generally discussed in the absence of the world of culture (world 3). In sharp contrast, Popper (1982) has taken world 3 as an indispensable part of the evolution of consciousness or minds (world 2). Hominin minds did not just produce such stone tools but had also been affected and transformed by those tools—by their own creations. We are creatures of our own making, since world 2 not only creates world 3, but evolves together and in interaction with the objective, extra-somatic, world 3 products (Popper and Eccles 1977, 442). This insightful observation of Popper enables us to see why the archaeological data of prehistoric stone tools—possibly the earliest creative (world 3) products of the (now extinct) hominin conscious-cognitive faculties—are of great significance for any critical-philosophical account of the origin and evolution of (hominin) minds or consciousness (world 2).

References

Arhem, P., Liljenstrom, H.: On the coevolution of consciousness and cognition. J. Theor. Biol. **187**, 601–612 (1997)

Averill, E., Keating, B.F.: Does interactionism violate a law of classical physics? Mind **90**(357), 102–107 (1981)

De Beaune, S.A.: The invention of technology: prehistory and cognition. Curr. Anthropol. **45**(2), 139–162 (2004)

Bondi, H.: The philosopher for science. Nature **358**, 363 (1992)

Campbell, D.T.: Downward causation in hierarchically organized biological systems. In: Ayala, F.J., Dobzhansky, T. (eds.) Studies in the Philosophy of Biology, pp. 179–186. Macmillan, London (1974)

Churchland, P.: Matter and consciousness: a contemporary introduction to the philosophy of mind. MIT Press, Cambridge MA (1984)

Clark, A., Chalmers, D.: The extended mind. Analysis **58**, 7–19 (1998)

Cornman, J.W.: A nonreductive identity thesis about mind and body. In: Feinberg, J. (ed.) Reason and Responsibility: Readings in Some Basic Problems of Philosophy. Dickenson Publishing Co., Encino (1978)

Davies, P.C.W.: The physics of downward causation. In.: Clayton, P., Davies, P. (eds.) The Re-Emergence of Emergence, pp. 35–51. Oxford University Press, Oxford (2006a)

Davies, P.C.W.: Preface. In.: Clayton, P., Davies, P. (eds.) The Re-Emergence of Emergence, pp. ix–xiv. Oxford University Press, Oxford (2006b)

Dennett, D.C.: Current issues in the philosophy of mind. Am. Philos. Q. **15**(4), 249–261 (1978)

Dennett, D.C.: Review of the self and its brain: an argument for interactionism. J. Philosophy. **76**(2), 91–97 (1979)

Dennett, D.C.: Granny versus mother nature—no contest. Mind Lang. **11**(3), 263–269 (1996)

Ellis, G. F. R.: On the nature of emergent reality. In.: Clayton, P., Davies, P. (eds.) The Re-Emergence of Emergence, pp. 79–107. Oxford University Press, Oxford (2006)

Fodor, J.: Deconstructing dennett's Darwin. Mind Lang. **11**(3), 246–262 (1996)

Greene, B.: The elegant universe. W.W. Norton & Company, New York (1999)

Heil, J.: Philosophy of mind: a contemporary introduction. Routledge, New York (1998)

Heil, J.: Philosophy of mind: a guide and anthology. Oxford University Press, Oxford (2004)

Hihara, S., et al.: Extension of corticocortical afferents into the anterior bank of the intraparietal sulcus by tool-use training in adult monkeys. Neuropsychol. **44**, 2636–2646 (2006)

James, W.: Are we automata? Mind **4**, 1–22 (1979)

Jeffares, B.: The co-evolution of tools and minds: cognition and material culture in the hominin lineage. Phenomenol. Cogn. Sci. **09**, 503–520 (2010)

Jones, M.W.: Electromagnetic-field theories of mind. J. Conscious. Stud. **20**(11–12), 1–26 (2013)

Jones, M.W.: How to make mind-brain relations clear. J. Conscious. Stud. **17**(5–6), 135–160 (2010)

Kim, J.: Being realistic about emergence. In. Clayton, P., Davies, P. (eds.) The Re-Emergence of Emergence, pp. 189–202. Oxford University Press, Oxford (2006a)

Kim, J.: Emergence: core ideas and issues. Synthese. **151**(3), 547-559 (2006b)

Libet, B.: A testable field theory of mind—brain interaction. J. Conscious. Stud. **1**(1), 119–126 (1994)

Libet, B.: Conscious mind as a field. J. Theor. Biol. **178**, 223–224 (1996)

Lindhal, B.I.B.: Estimating Popper's impact. Nature **360**, 204 (1992)

Lindhal, B.I.B.: Consciousness and biological evolution. J. Theor. Biol. **187**, 613–687 (1997)

Lindhal, B.I.B., Arhem, P.: Mind as a force field: comments on a new interactionistic hypothesis. J. Theor. Biol. **171**, 111–122 (1994)

Lindhal, B.I.B., Arhem, P.: Consciousness and neural force fields. J. Conscious. Stud. **23**(7–8), 228–253 (2016)

Lowe, E.J.: An introduction to the philosophy of mind. Cambridge University Press, Cambridge (2000a)

Lowe, E.J.: Causal closure principles and emergentism. Philosophy. **75**(294), 571–585 (2000b)

Malafouris, L.: The Brain-Artefact Interface (BAI): a challenge for archaeology and cultural neuroscience. Soc. Cogn. Affect. Neurosci. **5**, 264–273 (2010)

Malafouris, L.: How things shape the mind: a theory of material engagement. The MIT Press, Cambridge, MA (2013)

Malafouris, L.: Neuroarchaeology: exploring the links between neural and cultural plasticity. In: Chiao, J.Y. (ed.) Cultural Neuroscience: Cultural Influences on Brain Function, pp. 253–261. Elsevier, Amsterdam (2009)

Malafouris, L.: Metaplasticity and the primacy of material engagement. Time and Mind **8**(4), 351–371 (2015)

Mithen, S. (ed.): Creativity in human evolution and prehistory. Routledge, London (1998)

Mithen, S., Parsons, L.: The brain as a cultural artefact. Camb. Archaeol. J. **18**(5), 415–422 (2008)

Moore, M.W., Perston, Y.: Experimental insights into the cognitive significance of early stone tools. Plos One **11**(7), (2016). https://doi.org/10.1371/journal.pone.0150341

Nagel, T.: What it is like to be a Bat? Philos. Rev. **83**, 435–450 (1974)

Nowell, A., Davidson, I. (eds.): Stone tools and the evolution of human cognition. University Press of Colorado, Boulder (2010)

Orban, G.A., Caruana, F.: The neural basis of human tool use. Front. Psychol. **5**, 3–10 (2014)

Popper, K.R.: Natural selection and the emergence of the mind. Dialectica **32**(3–4), 339–355 (1978)

Popper, K.R.: Three worlds. Mich. Q. Rev. **18**(1), 1–23 (1979a)

Popper, K.R.: Objective knowledge: an evolutionary approach. Oxford University Press, Oxford (1979b)

Popper, K.R.: The open universe: an argument for indeterminism. Routledge, London (1988)

Popper, K.R.: Knowledge and the body-mind problem. Routledge, London (1994)

Popper, K.R., Eccles, J.C.: The self and its brain. Springer Verlag, Berlin (1977)

Popper, K.R.: Scientific reduction and the essential incompleteness of all science. In: Ayala, F. J., Dobzhansky, T. (eds.) Studies in the Philosophy of Biology, pp. 259–284. Macmillan, London (1974)

Popper, K.R.: The place of mind in nature. In: Elvee, R.Q. (ed.) Mind in Nature, pp. 31–59. Harper & Row, Cambridge (1982)

Popper, K.R.: Critical remarks on the knowledge of lower and higher organisms, the so-called sensory motor systems. In.: Creutzfeldt, O., Schmidt, R.F., Willis, W.D. (eds.) Sensory-Motor Integration in the Nervous System, pp. 19–31. Springer Verlag, Berlin (1984a)

Popper, K.R.: Evolutionary epistemology. In.: Pollard, J.W. (ed.) Evolutionary Theory: Paths into the Future, pp. 239–255. John Wiley & Sons, New York (1984b)

Popper, K.R., Lindhal, B.I.B., Arhem, P.: A discussion of the mind-brain problem. Theor. Med. **14**, 167–180 (1993)

Ronan, C.A.: The Cambridge illustrated history of the world's science. Cambridge University Press, Cambridge (1983)

Rooijen, J.V.: Interaction and evolution: a critique of Popper. Br. J. Philos. Sci. **38**(1), 87–92 (1987)

Rupert, R.D.: Challenges to the hypothesis of extended cognition. J. Philosophy. **101**(8), 389–428 (2004)

Settle, T.: Van Rooijen and Mayr versus Popper: is the Universe causally closed? Br. J. Philos. Sci. **40**(3), 389–403 (1989)

Skoyles, J.R.: Popper: Success or failure? Nature **359**, 100 (1992)

Stout, D., Chaminade, T.: The evolutionary neuroscience of tool making. Neuropsychol. **45**, 1091–1100 (2007)

Stout, D., Chaminade, T.: Making tools and making sense: complex, intentional behaviour in human evolution. Camb. Archaeol. Journal. **19**(1), 85–96 (2009)

Stout, D., Toth, N., Schick, K., Chaminade, T.: Neural correlates of early stone age tool making: technology, language and cognition in human evolution. Philos. Trans. R. Soc. B **363**, 1939–1949 (2008)

Stout, D.: Possible relations between language and technology in human evolution. In.: Nowell, A., Davidson, I. (eds.) Stone Tools and the Evolution of Human Cognition, pp. 159–184. University Press of Colorado, Boulder (2010)

Stout, D.: Stone tool making and the evolution of human culture and cognition. Philosophical Transactions of the Royal Society of London. Series B 366, 1050–1059 (2011)

Stout, D., Chaminade, T.: Stone tools, language and the brain in human evolution. Philosophical Transactions of the Royal Society of London. Series B, 367, 75–87 (2012)

Taylor, E.: An explication of emergence. Philos. Stud. **172**(3), 653–669 (2015)

Tennie, C. et. al.: Early stone tools and cultural transmission: resetting the null hypothesis. Curr. Anthropol. **58**(5), 652–672 (2017)

De la Torre, I.: The origins of stone tool technology in africa: a historical perspective. Philos. Trans. R. Soc. Lond. B: Biol. Sciences. **366**(1567), 1028–1037 (2011)

Toth, N.: The first technology. Sci. Am. **1**, 112–121 (1987)

Toth, N., Schick, K.D.: The Oldowan: the tool making of early hominins and chimpanzees compared. Ann. Rev. Anthropol. **38**, 289–305 (2009)

Van Gulick, Robert.: Reduction, emergence and other recent options on the mind-body problem: a philosophic overview. J. Conscious. Stud. **8**(9–10), 1–34 (2001)

Washburn, S.L.: The evolution of man. Sci. Am. **239**(3), 194–211 (1978)

Weinberg, S.: Dreams of a final theory. Vintage Books (1994)

Wynn, T.: Hafted spears and the archaeology of mind. Proc. Natl. Acad. Sci. **106**(24), 9544–9545 (2009)

Wynn, T.G., et al.: An ape's view of the Oldowan revisited. Evol. Anthropol. **20**(5), 181–197 (2011)

Wynn, T., Coolidge, F.L.: Archaeological insights into hominin cognitive evolution. Evol. Anthropol.: Issues, News, Reviews. **25**(4), 200–213 (2016)

Wynn, T.G., Coolidge, F.L. (eds.): Cognitive models in palaeolithic archaeology. Oxford University Press, Oxford (2017)

Manjari Chakrabarty is Associate Professor in Philosophy at Visva Bharati (Central University), Santiniketan, India. Her current research focuses mainly on two fields: on Karl Popper's work (in the fields of epistemology, philosophy of science, and metaphysics), and on the philosophy of technology (most importantly on prehistoric stone-tool technology, on the philosophy of human-artefact relations and on the epistemic dimension of technological artefacts). She is presently investigating how archaeological findings on stone-tool production and cognitive evolution could be illuminated by insights drawn from Karl Popper's philosophical writings. Her key articles include "How Stone Tools Shaped Us: Post-Phenomenology and Material Engagement Theory" in *Philosophy and Technology* (2018), "A Philosophical Study of Human-Artefact Interaction" in *AI &Society: Journal of Knowledge, Culture and Communication (2017),* "A Philosophical Inquiry into the Character of Material Artifacts", *Philosophia Scientiae (2014)*, "Karl Popper on Theory Creation" in *International Journal of Arts and Sciences* (2010). Email: manjari.chakraborty@visva-bharati.ac.in

Popper's Emergentism

Olga Markič

1 Introduction

Karl Popper is one of the most famous advocates of emergentism in the philosophy of mind. In *The Self and its Brain* (1977), which he co-wrote with neurophysiologist John Eccles, he argues for psychophysical interactionism and the emergence of consciousness. Their arguments for this position are based on his thesis of the creative evolution of the three worlds developed in *Objective Knowledge* (1972). He proceeds from two sides. One is constructive, in which he presents those three worlds, the theory of creative evolution and the emergence of new forms as building blocks for his solution to the mind-body problem. The other uses negation. He criticises four materialist/physicalist positions (Popper and Eccles 1977, 51–54). The first is radical physicalism (or radical behaviourism) which negates the existence of conscious or mental processes. This seems to Popper most implausible as it denies such undeniable facts as subjective pain. The three remaining positions—panpsychism, epiphenomenalism and identity theory—admit the existence of mental (conscious) processes but also accept the fundamental principle of physicalism, i.e. the closedness of World 1 which, as Popper argues, cannot hold. This claim is based on his epistemology and his three worlds. Thus I will first present Popper's three worlds, followed by a discussion of the theory of creative evolution and an analysis of the concept of emergence and of his claim about the strong relationship between novelty and unpredictability. I will analyse two different senses in which Popper uses the concept of emergence: (1) emergence as the unpredictability/nonderivability/nondeducibility of properties, and (2) emergence as the absolute unpredictability of events (cf. Stephan 1992, 34). This distinction parallels the distinction between the epistemological and ontological concepts of emergence. I will argue that Popper, despite some examples which point to an epistemological notion, suggests ontological emergentism with 'downward

---

O. Markič (✉)
Department of Philosophy, Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia
e-mail: olga.markic@ff.uni-lj.si

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_17

causation'. This enables him to see humans as autonomous and free. However, it is also vulnerable to the problems of interactionist dualism, particularly the question of its compatibility with some basic principles of physical science. I will show how Popper's position differs from the weaker emergentism present in various theories in cognitive science, such as the theories of emergent properties in connectionism and neural networks. I will conclude with some thoughts regarding neurophilosophy, a framework for collaborative research among philosophers and cognitive scientists.

2 Popper's Epistemology and His Three Worlds

Popper's proposal regarding the mind-body relation is closely connected to his views about epistemology and his notion of the three worlds (Popper 1972). Popper argues for a pluralistic philosophy whereby the world consists of at least three ontologically distinct sub-worlds: the physical world or world of physical states (World 1), the mental world, or world of mental states (World 2) and the world of intelligibles, or world of ideas (in the objective sense) (World 3) (Popper 1972, 1977). World 3 is the world of the products of the human mind, such as "stories, explanatory myths, tools, scientific theories (whether true or false), scientific problems, social institutions, and works of art" (Popper and Eccles 1977, 38). Although it is similar to Plato's world of Forms or Ideas, Popper's World 3 differs in being human-made. According to Popper, this does not mean that later on the theories could not live their own life. His standard example is from arithmetic. "A number system may be said to be the construction or invention of men rather than their discovery. But the difference between even and odd numbers, or divisible and prime numbers, is a discovery: these characteristic sets of numbers are there, objectively, once the number system exists, as the (unintended) consequences of constructing the system; and their properties may be discovered." (ibid., 40) He concludes that this and similar considerations establish "the objectivity of World 3, and its (partial) autonomy" (ibid., 41). This means that "we can get more out of World 3 than we ourselves put into it" (Popper 1994, 31). Popper suggests that his theory of World 3 leads to a view of human creativity in science that is closely related to evolutionary theory and of which more will be said in the next section.

Another important feature for which Popper argues is the reality of the three worlds. To understand his thesis about the reality of Worlds 2 and 3, we have to look at how Popper understands the term 'real'. For him, "the most central usage of the term 'real' is its use to characterize material things of ordinary size - things which a baby can handle and (preferably) put into his mouth. From this, the usage of the term 'real' is extended, first, to bigger things - things which are too big for us to handle, like railway trains, houses, mountains, the earth and the stars; and also to smaller things - things like dust particles or mites. It is further extended, of course, to liquids and then also to air, to gases and to molecules and atoms" (Popper and Eccles 1977, 9).

According to Popper, there is the following principle behind the extension: "the entities which we conjecture to be real should be able to exert a causal effect upon the *prima facie* real things; that is, upon material things of an ordinary size: that we can explain changes in the ordinary material world of things by the causal effects of entities conjectured to be real" (ibid.). Although Popper suggests solid material things are a paradigm of reality, according to this principle abstract entities are also real. In physics these are, for example, forces and fields. Although we can describe them as abstract theoretical entities, they directly or indirectly interact with ordinary material things and are thus real. So, for Popper, things are real "if they can causally act upon, or interact with, ordinary real material things" (ibid., 10).

It seems obvious that many World 3 objects are embodied in World 1, such as books, computers and paintings. They belong to both World 3 and World 1. Some, like poems, may also exist as memories and thus World 2 objects, also being encoded in human brains (World 1). Popper stresses that what makes a book an important product of the human mind is not the book as a World 1 object, but its content. It remains unchanged, although presented in different printings and editions. It is the content that belongs to World 3. Popper goes even further and argues that there are also unembodied World 3 objects which belong neither to World 1 nor to World 2. "It is important to realize that the objective and unembodied existence of these problems [like the previously mentioned problems of the number system—O.M.] precedes their conscious discovery in the same way as the existence of Mount Everest preceded its discovery; and it is important that the consciousness of the existence of these problems leads to the suspicion that there may exist, objectively, a way to their solution, and to the conscious search for this way" (ibid., 41–42). So, if unembodied World 3 objects exist, there has to be a way to grasp them "by a method which depends little, if at all, upon their embodiment or upon the use of our senses" (ibid., 43). In contrast with Plato, who refers to intellectual intuition and where grasping is a kind of seeing, Popper understands the grasping of a World 3 object as an active process, and explains understanding World 3 objects in terms of making or re-making them (ibid., 44).

The decisive role in connecting the three worlds is, according to Popper, given to consciousness. Consciousness enables human beings to produce, understand and be influenced by abstract objects such as theories, norms and values. These abstract objects (World 3) can be grasped through human mental processes (World 2) and thereby influence the physical world (World 1). World 2 is a kind of mediator between World 1 and World 3 in which the mind establishes an indirect link between them. Popper emphasizes the importance of learning to use language and thus to grasp objective thought concepts. He suggests that such considerations support the objective reality of all three worlds and thus also the existence of a subjective mental world of personal experiences (Popper 1972, 155–156). Therefore, interaction with World 3 takes the mental world beyond the physical world and thus supports the interactionist solution to the mind-body problem.

3 Creative Evolution and the Emergence of New Forms

Popper suggests that consciousness is a novelty which emerges from the basic structure of physical states and is able to interact both with the physical world and with the world of abstract objects. In contrast with the view, "There is nothing new under the sun" (Ecclesiastes 1:9), which he sees behind both physicalism and panpsychism, he argues that the universe is creative and that consciousness is an emergent phenomenon (Popper and Eccles 1977, 14).

Although Popper shares with physicalism both the thesis that material objects are the paradigms of reality, and the evolutionary hypothesis, he strongly opposes reductionism and criticises the layered system based on reductionists ideas as presented in the table below (ibid., 17):

(12) Level of ecosystems
(11) Level of populations of metazoa and plants
(10) Level of metazoa and multicellular plants
(9) Level of tissues and organs (and of sponges?)
(8) Level of populations of unicellular organisms
(7) Level of cells and of unicellular organisms
(6) Level of organelles (and perhaps of viruses)
(5) Liquids and solids (crystals)
(4) Molecules
(3) Atoms
(2) Elementary particles
(1) Sub-elementary particles
(0) Unknown: sub-sub-elementary particles?

He agrees with the reductionists' claim that the universe is layered (beginning with the sub-elementary particles and continuing with atoms, molecules, cells, organisms and ecosystems) but he denies that events and things on a higher level could be explained in terms of the lower levels. He provides two main critiques. The first concerns the organisation of the table, which should be much more complex, and stresses that even if one could present a better organized table there would still be difficulties because "the biosystem is not intrinsically organized in a neat stepwise hierarchy" (ibid., 17). This observation then leads to the second, crucial downside of reductionism and its idea that "what happens to a whole can be explained by way of explaining the structure (the arrangement) and the interaction of its parts" (ibid., 18).

In contrast to the reductionist programme he believes that there are stages in the evolution of the universe and that some of them are unpredictable or emergent. He arranges some of the cosmic evolutionary stages in the following table (ibid., 16):

World 3 (the products of the human mind)	(6) Works of Art and science (including technology). (5) Human Language. Theories of Self and of Death.
World 2 (the world of subjective experience)	(4) Consciousness of Self and of Death. (3) Sentience (Animal Consciousness).
World 1 (the world of physical objects)	(2) Living Organisms. (1) The Heavier Elements. Liquids and Crystals. (0) Hydrogen and Helium.

Popper argues that Darwin's theory of evolution provides the theoretical tools for such a viewpoint and that science suggests a picture of a universe that is inventive or even creative, and in which new things emerge, on new levels. "The evolution of the universe, and especially the evolution of life on earth—has produced new things: *real novelty* …. evolution has produced much that was not foreseeable, at least not for human knowledge" (ibid., 14–15). Real novelty and unpredictability are two of the main characteristics of creative or emergent evolution. In order to emphasise the importance of evolutionary changes that begin with new patterns of behaviour, he uses the term 'organic evolution', which was first used by British emergentist C. Lloyd Morgan (1923). Popper agrees with Darwin that either habits change first and structures afterwards, or slight modifications of structure lead to changed habits, and that in both cases it is natural selection which works on the genetic structure. But he also points out that "in many cases, and in some of the most interesting cases, habits change first" (Popper and Eccles 1977, 13).

He also suggests that by adopting a new form of behaviour the individual organism may also change its environment and that by doing so consciously, this may even lead to the construction of a new ecological niche by the organism (ibid., 12).

Thus, according to Popper, "the theory of organic evolution makes it understandable that the mechanism of natural selection becomes more efficient when there is a greater behavioural repertoire available. Thus it shows the selective value of a certain innate behavioural freedom - as opposed to behavioural rigidity which must make it more difficult for natural selection to produce new adaptations." (ibid., 13) He argues that even an animal is an active, problem solving *agent*, always actively attempting to control its environment (ibid., 127).

Consciousness emerges out of four biological functions: pain, pleasure, expectation and attention (ibid., 127). The crucial point is that evolution produces the mind and human language and the human mind produces tools, stories, explanatory myths and works of art and science. "We could say that in choosing to speak, and to take interest in speech, man has chosen to evolve his brain and his mind; that language, once created, exerted the selection pressure under which emerged the human brain and the consciousness of self." (ibid., 13) He thus suggests that if agent's choices, thoughts, plans and actions have repercussions on the agent, including the evolution of the brain, then these considerations also support the existence of conscious experience. He sees great problems for all of those who deny the existence of consciousness or who, although admitting that consciousness exists, claim that the physical world is causally closed (ibid., 14).

It is clear that Popper wants to preserve a special status for human consciousness. He sees consciousness as real, causally efficacious and irreducible to any lower level of existence. He suggests emergence as a solution to the mind-body problem which avoids two extremes—physicalist reductionism and Descartes' substance dualism. In contrast with the latter he argues that consciousness is a product of evolution and is not present from the beginning of the universe as a separate substance. In contrast with the former he points to the reality and non-reducibility of consciousness. He thus proposes emergentist interactionist dualism as the best solution to the mind-body problem, suggesting that consciousness is a novel structure that emerges from the basic structure of physical states and is able to interact with it.

His proposal closely resembles the approaches of the British emergentists (Markič 2004). Their motivation to employ the notion of emergent properties for mental properties was to preserve our common-sense belief that mental properties have causal powers and at the same time escape the difficulties of interactionist substance dualism. They believed that there were no nonphysical substances that were outside of the framework of space-time. However, while committing themselves only to the existence of concrete physical particles and their aggregates (as do physicalists), British emergentists also maintained that some physical things, i.e. humans, possess some special properties—mental properties that are novel and neither explicable nor deducible from physical properties, and that have genuine causal power (Markič 2004). It is interesting that Popper in *The Self and its Brain* mentions G. H. Lewes and his introduction of the term 'emerging' (Popper and Eccles 1977, 14) and C. L. Morgan in connection with 'organic evolution' (ibid., 12), but not S. Alexander and C. D. Broad, who are regarded as important representatives of traditional British emergentism (McLaughlin 1992; Stephan 1992).

Let us now look in more detail at how Popper uses the concept of emergence. I will first analyse how Popper relates novelty, unpredictability and explainability; then, I will look at the notion of downward causation.

3.1 Novelty and Unpredictability

As previously stated, Popper sees his position as being in opposition to the programme of physicalist reductionism and what he sees as its basic—yet false—intuition, namely "if the universe consists of atoms or elementary particles, so that all things are structures of such particles, then every event in the universe ought to be explicable, and in principle predictable, in terms of *particle structure* and of *particle interaction*" (Popper and Eccles 1977, 17).

Traditionally, a reductionist description is a view that the whole can be reduced to its parts, i.e. the property of the whole is a mere aggregate of its parts' properties, whilst in contrast, emergent properties, the properties of the whole system, are more than just the sum of its parts' properties. They are not merely 'resultant', i.e. additive combinations of components (Markič 2004). In contrast with reductionist systems, in the course of evolution 'genuine novelties' arise from pre-existing building blocks. These novelties are new structures that constitute new entities with new properties. Popper gives an example:

> there is the fact that in a universe in which there once existed (according to our present theories) no elements other than, say, hydrogen and helium, no theorist who knew the physical laws then operative and exemplified in this universe could have predicted all the properties of the heavier elements not yet emerged, or that they would emerge; or all the properties of even the simplest compound molecules, such as water. (Popper and Eccles 1977, 11)

Emergent properties are thus novel, and also unpredictable from the basic properties and the laws concerning the parts of the system.

It seems that Popper understands the term 'unpredictable' as equivalent to 'not derivable' or 'not deductively explainable'. When he criticizes atomists and their view that there should be no novelty except the novelty of arrangement, he writes,

> new atomic arrangements may lead to physical and chemical properties which are not derivable from a statement describing the arrangement of the atoms, combined with a statement of atomic theory. Admittedly, some such properties have been successfully derived from physical theory, and these derivations are highly impressive; yet it seems that the number and complexity of both the different molecules and their properties are unlimited and that they may far transcend the possibilities of deductive explanation. (ibid., 23)

The above passages suggest that one can understand emergence as an epistemological thesis of both practical unpredictability and the impossibility of deductive explanation. However, I wonder if this sense of emergence is strong enough to support Popper's ideas about consciousness as causally efficacious. Stephan (1992, 35) is even more critical in his analysis of Popper's concept of emergence as the unpredictability/nonderivability/nondeducibilty of properties. He thinks that it is much too broad: from such a perspective nearly all properties would be emergent. One would even be unable to deduce the weight of a whole from the weights of its parts. To do this, one must invoke the principle of the additivity of weight, and this principle, whilst nomological, is logically contingent. So, even the weight of a whole would be an emergent property of the whole (Stephan 1992, 35).

Popper is probably aware that beside the concept of emergence as the unpredictability/nonderivability/nondeducibilty of properties, he also needs a stronger notion of emergence. This second sense—the absolute unpredictability of events—is based upon an opposition to determinism (Markič 2004; Stephan 1992). Popper quotes the famous Laplace formulation:

> We ought … to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Assume … an intelligence which could know all the forces by which nature is animated, and the states at an instant of all the objects that compose it; … for [this intelligence], nothing would be uncertain; and the future, as the past, would

> be present to its eyes. If this Laplacean determinism is accepted, nothing whatever can be unpredictable in principle. So evolution cannot be emergent. (Popper and Eccles 1977, 22)

Popper points out that in the above passage Laplace is not concerned with objectively indeterminate or chance-like events but only with events of which humans have subjectively insufficient knowledge. In Laplace's time, physicists believed that the physical world was deterministic. However, as Popper emphasises, things have changed with the modern physics which assumes "that there are objectively chancelike events, and objective probabilities or propensities." (ibid., 23) He argues that the findings of modern physics, i.e. that the interaction between atoms has a random aspect, are enough "to destroy the old atomistic intuition of a mechanical determinism" (ibid., 34).

This second sense of emergence is based on a thoroughgoing indeterminism and the change from Newtonian physics to quantum physics. Popper's conclusion is that "the emergence of hierarchical levels or layers, and of an interaction between them, depends upon a fundamental indeterminism of the physical universe. Each level is open to causal influences coming from lower and from higher levels" (ibid., 35).

It seems to me that Popper puts too much weight on indeterminism at the quantum level. It is questionable whether event indeterminism on the lower level (quantum physics) really makes any difference concerning the nature of mental properties and the relation between mental and physical properties. The question is still open and some authors (e.g. Penrose 1989, 1994; Hameroff and Penrose 1996) argue that quantum theory applies to consciousness. However, as discuss in the next section, indeterminism at the quantum level is not the only source of indeterminism. To explain indeterminate processes at higher levels, Popper uses the theory of evolution and the process of natural selection. These are even more important for understanding why mind and consciousness are not totally determined and can be free.

Popper's usage of emergence as novelty and unpredictability is not as precise as one may wish. There can still be doubts regarding whether it is possible to understand emergentism as an epistemological thesis of practical unpredictability, or whether new emergent properties are unpredictable in principle and carry ontological consequences. In the next section I show similar difficulties with understanding the concept of downward causation. Some examples suggest a weaker notion of emergentism but Popper wants to propose a stronger version where consciousness has its own independent causal power.

3.2 Downward Causation

In his discussion of emergence Popper refers to downward causation whenever a higher structure operates causally upon its substructure. He admits that it is much easier to understand upward causation, how the substructures of a system cooperate to affect the whole system. Yet because the set of substructures interacts causally in any case, it seems there is no room for any interference upon an action from above (Popper

1978, 348). He gives some physical examples, such as lasers and holograms, and gives the following explanation: "the whole, the macro structure, may *qua* whole, act upon a photon or an elementary particle or an atom….And there are also many other macro structures which are examples of downward causation: every simple arrangement of negative feedback, such as a steam engine governor, is a macroscopic structure that regulates lower level events, such as the flow of the molecules that constitute the steam" (ibid., 19). It is interesting that in the above characterization and examples, Popper uses the terms "act upon" and "regulates" instead of "causation". It is not clear what exactly he means by them and one can interpret it as some specific operation through molecular interaction. He continues that the most interesting examples of downward causation are to be found in organisms and in societies of organisms. He gives an example of an animal which may survive the removal of an organ and the death of many of its cells, but when the animal dies, the death will lead to the death of its constituent parts, its cells (ibid., 19).

With these examples he supports his critique of those reductionists who deny the possibility of any sort of downward causation, but it is not clear that these examples violate the closure of the physical world. I suggest they could be characterized as a form of weak emergentism. In contrast, strong emergentism postulates new causal powers with ontological consequences (Markič 2004; Stephan 2002). Popper uses this strong, causal emergentism when he talks about consciousness and mental phenomena. I concur with Tim Crane's emphasis upon downward causation being one of the main lures of emergentism, one consequence being a denial of the causal closure of physics: "not *all* physical effects are entirely fixed by purely physical causes: in some cases, mental states are needed as well" (Crane 2001, 63). As Popper states, "even those who think that mind is 'just' the causal product of self-organizing matter should feel that it is difficult to regard the Ninth Symphony in this way, or Othello, or the theory of gravitation." (Popper and Eccles 1977, 207) I will return in the next section to the discussion of the causal closure of the physical world, but let us look first at Popper's consideration of natural selection as a means to a better understanding of emergence.

Popper is well aware of Thomas Henry Huxley's proposal in "On the hypothesis that animals are automata, and its history" (1874), where Huxley accepts the existence of subjective experiences but denies that they can have any effect upon the machinery of the human body or brain. Popper criticises Huxley and thinks that such a solution is a consequence of the scientific framework, i.e. physical determinism, which Huxley accepts. As Popper observes, in this way Huxley, a friend and great supporter of Darwin, proposes a solution which seems implausible from the standpoint of the theory of evolution. Namely, Darwin discusses mental powers in animals and humans and argues that they are a product of natural selection. So, mental powers must help in the struggle for physical survival and are thus able to exert an important influence on the physical actions of animals and men. Thus, Popper concludes, animals and men could not be automata in Huxley's sense (Popper 1978, 350).

Popper emphasises the role of random processes and indeterminism in understanding natural selection and consequently in downward causation. He acknowledges that Darwin worried he could not explain variation and was uneasy to take it as chance-like, but he discovers that "downward causation can sometimes at least be explained as selection operating on the randomly fluctuating elementary particles. The randomness of the movements of the elementary particles - often called 'molecular chaos' - provides, as it were, the opening for the higher-level structure to interfere. A random movement is accepted when it fits into the higher level structure; otherwise it is rejected" (ibid., 348).

As mentioned at the end of the previous section, it is not clear how quantum indeterminacy can help us understand how this indeterminacy leads to human action and not just to random movements. Popper is aware of such difficulties but tries to provide plausible upgrades. He thinks that "a choice process may be a selection process, and the selection may be from some repertoire of random events, without being random in its turn." (ibid., 349) In this way, an act of choice may even be seen as an act of free will. He supports his claim by analogy with genetic mutations, which seem to be brought about by quantum theoretical indeterminacy. They are probabilistic and not in themselves originally selected or adequate, but on them then operates natural selection which eliminates inappropriate mutations. Popper argues that a similar process goes on with respect to new ideas and to free-will decisions. There are possibilities brought about by the brain—a probabilistic and quantum mechanically characterized set of proposals, and then a kind of selective procedure eliminates those proposals and those possibilities which are not acceptable to the mind, anchored in World 3. The mind tries them out in World 3 and checks them by World 3 standards (Popper and Eccles 1977, 540). Popper is aware that this is just an analogy but he thinks it is at least worth speculating about such a proposal.

4 Weak and Strong Emergentism and the Causal Closure of the Physical Domain

I have shown how Popper's usage of the concept of emergence is sometimes ambiguous but nevertheless, when used in connection with consciousness, it is understood as strong emergence, with new causal powers at the emerging level. Let me first characterize weak and strong emergence and then continue with the discussion on the thesis of causal closure.

I have already referred to the British emergentists who discussed emergence at the end of the nineteenth and beginning of the twentieth centuries. Almost a hundred years later, emergence is receiving new attention owing to its use in descriptions of complex systems, such as connectionist models, neural networks and dynamical systems (e.g. Bechtel and Abrahamsen 2002). Cognitive scientists build models by running computer simulations in which the behaviour of the system emerges from interactions between simple components. In contrast, in the cognitivist approach to cognitive modelling, representations are manipulated according to certain rules. Functional decomposition is the method used by cognitivists and so it is clear that the resultant properties can be predicted and explained from the properties of the components. However, in modelling complex systems one cannot simply derive the result from the components but must run the simulation.

According to Mark Bedau (2008) this lack of simple derivation and the need to run simulations in order to gain results marks such systems as weakly emergent. Bedau characterizes weak emergence as follows: "The system's global behavior derives just from the operation of micro-level processes, but the micro-level interactions are interwoven in such a complicated network that the global behavior has no simple explanation. The central idea behind weak emergence is that emergent causal powers can be derived from micro-level information but only in a certain complex way." (Bedau 2008, 160) In contrast with weak emergence, in strong emergence the emergent properties are irreducible causal powers which have effects at both the macro and the micro level. Bedau adds a further requirement pertaining to the micro determination of the macro, claiming that emergent properties are supervenient on the basic properties (ibid., 158–159). The crucial distinction between weak and strong emergence is downward causation, which is supported by the strong version of emergence, but not the weak.

Scientists exploring complex systems mostly accept weak emergence as a satisfactory framework. With new mathematical tools they are able to model self-organizing processes with feedback, processes that are not easily decomposable. In the sense in which reduction means decomposition, such systems are not reductionistic. However, sometimes reduction is viewed in a broader sense, not in the old-fashioned sense of reduction, but more as the possibility to model and simulate macro processes on micro parts. What counts as macro or micro depends upon the context of the investigation.

Weak emergence also seems attractive for non-reductive physicalists, who want to preserve an active role for the mental, yet at the same time remain metaphysical physicalists. To secure the latter it is generally accepted that one need subscribe to the principle of physical causal closure: "If you pick any physical event and trace out its causal ancestry or posterity, that will never take you outside the physical domain. That is, no causal chain will ever cross the boundary between the physical and non-physical" (Kim 1998, 40).

However, Kim (1998) constructs an argument that attempts to demonstrate that non-reductive physicalism is an inconsistent position. With the help of the exclusion argument he further claims that anyone who accepts mind-body supervenience and the causal closure principle must also embrace reductionist identity theory. "Given that every physical event that has a cause has a physical cause, how is a mental cause also possible?" (Kim 1998, 38) The principle of causal exclusion states: If an event *e* has a sufficient cause *c* at *t*, no event at *t* distinct from *c* can be a cause of *e* (unless this is a genuine case of causal overdetermination). He suggests that it seems most unlikely that we might find cases in which mental properties and physical properties causally overdetermine the same effects in a systematic way. Nevertheless, many discussions revolve around this question and many authors believe that this is the best response to Kim's exclusion argument (Roche 2014).

Almost nobody in this debate questions the principle of causal closure, mainly because it seems that from the negation of the principle, an incompatibility with science ensues. Kim argues, "If you reject this principle, you are ipso facto rejecting the in-principle completability of physics, that is, the possibility of a complete and comprehensive physical theory of all physical phenomena." (Kim 1998, 38) Bedau, commenting on strong emergence, writes:

> There is no evidence that strong emergence plays any role in contemporary science. The scientific irrelevance of strong emergence is easy to understand, given that strong emergent causal powers must be brute natural phenomena. Even if there were such causal powers, they could at best play a primitive role in science. Strong emergence starts where scientific explanation ends. (Bedau 2008)

Lei Zhong (2019) has recently argued that we have to rethink the status of the closure principle. By examining exclusion arguments and possible solutions he concludes that "neither physical evidence nor physicalist considerations can satisfactorily support Closure" (ibid., 14).

Popper and Eccles (1977) seek compatibility with science and do not postulate any supernatural entities. They develop their theory with consciousness as real and causally potent, as a product of evolution. Minds, human language, works of art and of science,

> all this, so it seems, has evolved without any violation of the laws of physics. But with life, even with low forms of life, problem-solving enters the universe; and with the higher forms, purposes and aims, consciously pursued. We can only wonder that matter can thus transcend itself, by producing mind, purpose, and a world of the products of the human mind. (ibid., 11)

Do we have to negate the causal closure of the physical world to understand the whole picture? Perhaps we do not yet know enough and can only speculate upon potential scientific revolutions.

5 Conclusion

Popper's objective epistemology and three ontologically distinct worlds—in which the middle World 2, the world of subjective experience, interacts with both the physical World 1 and with World 3, the world of the products of the human mind—together with the theory of creative evolution and the emergence of consciousness, leads to an interactionist dualistic position upon the mind-body problem. At first glance it seems very similar to Descartes' substance dualism, in which mind and body also interact causally. The position is challenged from two main directions. One concerns the question of compatibility with some basic principles of physical science, as briefly discussed above. The other concerns the question how the nonextended and immaterial mind can move the body, an issue most notably raised by Princess Elizabeth of Bohemia. Descartes himself speculated that the pineal gland is "the part of the body in which the soul directly exercises its functions" (Descartes, in Lokhorst 2020). This answer is unsatisfactory by present scientific standards. However, as Descartes is known to have been both an excellent scientist and exceedingly interested in anatomy and physiology, Watson suggests that "if Descartes were alive today, he would be in charge of the CAT and PET scan machines in a major research hospital" (Watson, in Lokhorst 2020) and in this way would be able to substantially upgrade his proposal.

In a way, *The Self and its Brain* continues looking for answers to the challenges to Descartes' solution. Popper's discussions on the emergence of consciousness, its relationship with brain processes, and the collaboration with his colleague John Eccles, a Nobel Prize-winning neurophysiologist, in fact marks the beginning of the modern era of collaborative research between philosophers and neuroscientists. Patricia Churchland later wrote *Neurophilosophy* (1986) and thereby coined the new term for this interdisciplinary research programme. Her ideas about the relation between mind and body are physicalistic, sometimes more eliminativist and sometimes more reductionist. This is the reason why the term 'neurophilosophy' is often associated with these two positions. Thus, it may look strange that the birth of neurophilosophical research was, in fact, marked by emergentist interactionist dualism and Eccles' attempt to discover how and where the causal connection between the conscious self and the brain occurs. I propose that these differences show that there is no single 'right' metaphysical position in neurophilosophy. As Henrik Walter points out, it is best viewed "as a discipline that moves in on the mind-brain problem from two opposite directions. Either we begin on the empirical side and happen upon philosophical questions, or we set out with philosophical puzzles and need empirical findings to solve them.... It is best understood as a bridge discipline between subjective experience, philosophical theorizing, and empirical research." (Walter 2001, 125) To do this, it is good to bring both disciplines closer together without first precisely specifying the relation between the mental and the physical. Walter proposes three core theses of minimal neurophilosophy:

1. Ontology: mental processes of biological organisms are realized by or with the aid of neuronal processes.
2. Constraint: the philosophical analysis of mental processes should not contradict the best currently available brain theories.
3. Heuristic Principle: knowledge about the structure and dynamics of mental processes can be gained from knowledge about the structure and dynamics of neuronal processes (ibid., 132).

Within the framework of neurophilosophy, both disciplines do mutually affect one another. This can lead to the revision of some commonplace psychological notions. As Walter frames it, "In this way neurophilosophy has the potential to change our world view." (ibid., 126) Furthermore, Popper could add that World 3 theories always have the power to change us and our society, so we must also pay attention to their ethical and social implications.

Because interactionist dualism faces serious charges of incompatibility with science, Popper's emergent dualism is not popular among those philosophers who are more scientifically oriented. Even among philosophers who discuss emergentism, he is only rarely mentioned (e.g. Bedau and Humphreys 2008; McDonald and McDonald 2010). The debates in the philosophy of mind revolve predominantly around non-reductive physicalism and supervenience, whilst post-cognitivist approaches in the philosophy of cognitive science look for their inspiration more to phenomenology and Buddhism (e.g. Varela et al. 1991; Thompson 2007). However, at least some of his ideas, for example a feedback with World 3, creative and active engagement with the natural and social environment and the construction of a new ecological niche, may be interesting to explore in connection with the embodied approaches to cognition.

Popper's three worlds and his theory of the emergence of consciousness remind us that we have to be careful not to be too quick to accept the results of reductionist science as the final arbiter. We have to explicate our conjectures and carefully test them, both in science and philosophy. Thus, in conclusion, I quote from Popper's Darwin Lecture, which he delivered in 1977 at Darwin College, Cambridge. I believe it is even more relevant today than it was forty years ago. Nowadays, a surprising number of people are ready to believe conspiracy theories and are sceptical about science:

> The Darwinian revolution is still proceeding. But now we are also in the midst of a counter revolution, a strong reaction against science and against rationality. I feel that it is necessary to take sides in this issue, if only briefly; and also, in a Darwin lecture, to indicate where Darwin himself stood. My position, very briefly, is this. I am on the side of science and of rationality, but I am against those exaggerated claims for science that have sometimes been, rightly, denounced as 'scientism'. I am on the side of the search for truth, and of intellectual daring in the search for truth; but I am against intellectual arrogance, and especially against the misconceived claim that we have the truth in our pockets, or that we can approach certainty. (Popper 1978, 341)

References

Bechtel, W., Abrahamsen, A.: Connectionism and the mind. Blackwell, Oxford (2002)

Bedau, M.: Downward causation and autonomy in weak emergence. In: Bedau, M., Humphreys, P. (eds.) Emergence: Contemporary Readings in Philosophy and Science, pp. 155–188. MIT Press, Cambridge, MA and London (2008)

Bedau, M., Humphreys, P. (eds.) Emergence: contemporary readings in philosophy and science. MIT Press, Cambridge, MA and London (2008)

Churchland, P.S.: Neurophilosophy. MIT Press, Cambridge, MA (1986)

Crane, T.: Elements of mind: an introduction to the philosophy of mind. Oxford University Press, Oxford (2001)

Hameroff, S.R., Penrose, R.: Conscious events as orchestrated spacetime selections. J. Conscious. Stud. **3**(1), 36–53 (1996)

Kim, J.: Mind in a physical world. MIT Press, Cambridge, MA (1998)

Lokhorst, G.J.: Descartes and the pineal gland. The stanford encyclopedia of philosophy (Fall 2020 Edition). In: Zalta, E.N. (ed.) https://plato.stanford.edu/archivesfall2020/entries/pineal-gland/ (2020). Accessed 20 August 2020

Markič, O.: Causal emergentism. Acta Anal. **19**(33), 66–81 (2004)

McDonald, C., McDonald, G. (eds.): Emergence in mind. Oxford University Press, Oxford and New York (2010)

McLaughlin, B.: The rise and fall of British emergentism. In: Beckermann, A., Flohr, H., Kim, J. (eds.) Emergence or Reduction?, pp. 49–93. Walter de Gruyter, Berlin and New York (1992)

Morgan, C.L.: Emergent Evolution. Williams & Norgate, London (1923)
Penrose, R.: The emperor's new mind. Oxford University Press, Oxford (1989)
Penrose, R.: Shadows of the mind. Oxford University Press, Oxford (1994)
Roche, M.: Causal overdetermination and Kim's exclusion argument. philosophia (2014). https://doi.org/10.1007/s11406-014-9525-y
Popper, K.: Objective knowledge. Clarendon Press, Oxford (1972)
Popper, K.: Natural selection and the emergence mind. Dialectica **3**(4), 339–435 (1978)
Popper, K.: Knowledge and the body-mind problem. Routledge, London and New York (1994)
Popper, K., Eccles, J.: The self and its brain. Springer, New York (1977)
Stephan, A.: A systematic view on its historical facets. In: Beckermann, A., Flohr, H., Kim, J. (eds.) Emergence or Reduction?, pp. 25–48. Walter de Gruyter, Berlin, New York (1992)
Stephan, A.: Emergentism, irreducibility, and downward causation. Grazer Philos. Studien. **65**, 77–93 (2002)
Thompson, E.: Mind in life: biology, phenomenology, and sciences of the mind. Harvard University Press, Cambridge, MA and London (2007)
Varela, F., Thompson, E., Rosch, E.: The embodied mind. The MIT Press, Cambridge and London (1991)
Walter, H.: Neurophilosophy of free will: from libertarian illusions to a concept of natural autonomy. MIT Press, Cambridge, MA and London (2001)
Zhong, L.: Taking emergentism seriously. Australas. J. Philos. (2019). https://doi.org/10.1080/00048402.2019.1589547

Olga Markič is a Professor of Philosophy at the Faculty of Arts, University of Ljubljana, Slovenia, and at the Middle European interdisciplinary master programme in Cognitive Science (Mei:CogSci). She lectured as a visiting professor at different European universities and published papers and books in the area of logic and argumentation, philosophy of mind, philosophy of cognitive science and neuroethics (among them Cognitive science: Philosophical questions (2011), published in Slovenian). She is currently a President of the Slovenian Cognitive Science Society. E-mail: olga.markic@ff.uni-lj.si

The Place of the Mind in Nature

Joseph Agassi

The title of this essay, as of similar ones, is indebted to that of C. D. Broad's *The Mind and Its Place in Nature* of 1925. As it transpires, the title relates to the demise of the traditional theory of substance—a couple of decades earlier. Since the doctrine of the substance is the most significant idiosyncrasy of traditional western philosophy, its demise requires the radical rethinking of almost its entire heritage. The title of Broad indicates that the peculiarity of the mind invites a new study within the most general framework, the one we call "nature" that Spinoza has used as the most inclusive noun.

When speaking of nature, the default starting point may be historical. This would take the discussion straight to the oldest attitudes to nature. This may raise the question, of the oldest societies, which one embodies the most primitive philosophy. Fortunately, all primitive societies are similar: they are magically oriented and offer variants of the myth of creation. They have the same attitude to minds or rather to verbal meanings. Every system of thought that recognizes magic as effective seeks meaning in every event, thing, or system. It then allows dead people to refuse to leave the planet: they stay as ghosts. As ghosts represent their older selves, they tell us something about our living selves. What? How do the living and their ghosts merge? This is the oldest version of the mind-body problem. (In the present context, then, the words "ghost", "soul", "mind" and their likes are all synonymous).

Many philosophers find highly objectionable all systems that allow for magic. To distance themselves from it, they deny vehemently the existence of mental objects (those that in Popper's view inhabit the objective world of ideas, World3) as well as non-material objects such as the soul and as social institutions. This is not serious. Magic constitutes any anthropomorphic worldview in which intentions in the abstract fill all space.

J. Agassi (✉)
Tel Aviv University, Tel Aviv-Yafo, Israel
e-mail: agass@tauex.tau.ac.il

York University, Toronto, Canada

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_18

We have, then, to take it for granted that magic is not serious (together with all that surrounds it that goes this day by the name of the spiritual). Having given it up, we face a major difference between the living and the dead. The mere admission of this fact will not do, as mechanists declare the living body a smooth-running machine. This leaves out the observed fact that living things have free will; the same admission of the observation of life, then, includes the admission that we exercise free choice; In *The Open Society and Its Enemies* Popper insists on the need to admit it as sheer common sense regardless of any metaphysics that we may advocate. This is a strong argument, met by an equally strong observation of some sort of servomechanism or another. I will return to this in the next paragraph.

A part of this argument we must attend to now: when speaking of the place of life in nature, the default starting point must be some preferred philosophical system. Let us return to history, then, but if we return to antiquity, we may stay there for too long. So, rather than return to antiquity, we may try to make do with early modern philosophy. The first of these philosophical systems is the one that of Descartes has offered. He took the theory of substance for granted, as he began developing his philosophy allegedly with taking nothing for granted. Now the theory of substance, developed by the Eleatic school of Thales, is what characterizes western metaphysics and thus generally western intellectual practices. The more advanced philosophers of the twentieth century have abandoned the theory of substance (under the impact of both Russell's logic and of Einstein's physics, both of 1905). This raises the question, what is the soul or the mind that is under discussion here? Descartes claimed that animals other than humans have no souls or minds, that their mental abilities are utterly mechanical. (He envisaged robots three centuries before they appeared!) Some of his followers, chiefly la Mettrie, strongly disagreed: they said the same applies to humans: we too are soulless machines. To repeat, the rejection of the theory of substance requires some revision of this theory if not its outright rejection.

When speaking of nature, the default starting point may then very well be natural science. This, however, suggests a more fundamental starting point: what is science? This question takes up a huge literature, and so taking it up leads to the utter neglect of our initial interest. Fortunately, we can make do with one convention that the Royal Society of London has adopted at its foundation; it is the only one that the scientific tradition adheres to with no exception: science requires the adoption of all and only observations on record in the scientific literature as reported repeatedly, not reported impossible to repeat, and declared repeatable. (It may be an error, and often it is; and when proper reports refute a received observation report, it remains in the books after it undergoes the required qualifications).

Although this convention is uncontested within science, most philosophers who study science ignore it, particularly the once so authoritative Vienna Circle, whose members debated the status of observation reports: Rudolf Carnap took them to be certain and Otto Neurath did not. Rather, he forbade using mental terms (which taboo is downright magical). Carl G. Hempel discussed the justification of a generalization of a singular observation statement by its many repetitions, ignoring the fact that science admits observations only as generalizations.

When speaking of nature, the default starting point may then very well be all repeated repeatable observations of the mental. We may then wish to characterize the mental. The existence of servomechanisms—robots—shows that the challenge is not simple. Many philosophers, from Rudolf Carnap to John Searle, offer a simple answer: languages that describe servomechanisms are extensional, whereas to discuss meanings we need intentional terms. Now this distinction belongs to Bertrand Russell. He made the distinction between intentional and extensional terms in 1905 in order to overcome a trouble within logic: the true statements “Scott is the author of Waverly” and “the king wanted to know who the author of Waverly is” imply the false statement “the king wanted to know who Scott is”. (Waverly was a novel that Scott published anonymously.) Therefore, Russell excluded words such as “want” and “know” from the vocabulary of his system of logic: they are intensional. Since the intensional is the not-extensional and the extensional is the domain of formal logic, this renders formal logic, in particular Russell’s system of logic, quite useless for the present discussion. This, however, is not to allow being illogical with impunity.

When speaking of nature, the default starting point may then very well be all repeated repeatable observations of the intensional. We can now argue that discussing the mental as divine is scientifically irrelevant. This raises two difficult questions, namely, what is divine, and what is scientifically relevant? These difficult questions take us to innumerable controversies. Fortunately, all we need to eliminate the divine from the current discussion is to show that it is in principle not repeatable. This is easy, since allowing the divine to be repeatable renders it the object of all science. This chimes well with the philosophies of Spinoza and of Einstein, to mention some conspicuous examples. The exclusion of the divine from science is uninformative and not excluded. There is no serious objection to it. Tradition repeatedly explains a singular event by reference to divine will. Excluding singular events from science bars this discussion as unscientific. This holds also for explaining a singular occurrence of a repeatable event by reference to divine will. Explaining a repeatable event by reference to divine will, however, is generalizable to all repeatable events and so to science as such. This is uninformative and so its admission is but a matter of personal taste.

Another way to exclude the divine from science is to observe that statements that explain events by references to divine will are too arbitrary. This is true, but too weak as an objection. For, as a form of scientific objection to a theory it is the welcome invitation to strengthen in order to make it empirically testable (rather than arbitrary). Not so when an explanation with reference to the divine: strengthening it that way renders it possible to omit from it the reference to the divine without reducing its empirical content. Hence, the reference to the divine in it is scientifically redundant. This is true, but hardly explicable. The passionately anti-religious members of the Vienna Circle said, were it possible to offer testable assertions about the divine, they would not object to them. This is very generous of them, except that their generosity is redundant and the question before us is why is it impossible to offer testable hypotheses about the divine?

This seemingly very difficult question is easily answerable. Initially, references to the divine were testable. All magic views include a far-reaching hypothesis:

divine forces punish all crime. (See for example the classical Greek drama, such as *Oedipus* or *The Birds* or the biblical *Book of Judges.*) Whether this idea is empirically refutable or not is an open question. It seems, however, that some authorities deem it refuted (such as the complaint of Jeramiah about the triumph of evil, *Jeramiah* 12:1 and Plato's story of the ring of Gyges, *Republic*, 2:359a–2:360d). The idea that punishment reaches one after one's death makes it patently unempirical (as intended).

With the exclusion of the divine from scientific discourse, the discussion of the soul remains naturalist and thus hopefully scientific, namely, empirical. The wish to develop empirical science, then, should lead to listing some repeatable observations, seeking explanations of them that should be empirically testable, and testing them. This great idea William Whewell suggested clearly in the mid-nineteenth century. It was so bold that philosophers ignored it. It took another century before Karl Popper offered the breathtaking idea that the endorsement of this suffices for science to progress. Both left the process of the generation of hypotheses out of the discussion about science. The difference between Whewell and Popper is striking. Whewell wanted verification, and the weakness of his philosophy is merely in his failure to build a bridge between the corroboration and the verification of a theory. He stressed that an irrefutable hypothesis is unscientific, but he did not assert that a refutable hypothesis is scientific. This is where Popper came in, and this makes him a minimalist: that a scientific theory is refutable nowadays philosophers of science take for granted as a necessary condition for scientific character; most of them do not agree that refutability is also a sufficient condition for scientific character. It is not that they refuse to praise a hypothesis unless it is testable; it is that they do not agree with Popper's identification of the process of developing conjectures and refutations with scientific progress. Thus, representing the Establishment's view, Adolf Grünbaum, rejected Popper view as insufficient. As it happens, Grünbaum was a sworn atheist; Popper preferred to suspend judgment on this: he took naturalism as sufficient. What the present discussion shows is that naturalism is a part of the scientific tradition since the scientific revolution as well as a part of Popper's methodology.

One might object that not only Popper's methodology but also the convention of the scientific community plays a role in this discussion. Indeed, this is the case. Indeed, the whole of Popper's methodology is a series of proposals of conventions that direct the game of science. Seeing that science is a tradition, he rejected the naturalist view of it. (In this sentence naturalism is contrary not to supernaturalism but to conventionalism.) Whereas tradition deems convention arbitrary and conventionalism applies to the empirical content of theories, in Popper's philosophy it applies also to observation; it aims to maintain and increase openness to criticism and so it is the proposal to take scientific theories literally.

The question then is this. How do mind and body interact? Now they do in many ways, as the brain grows or suffers clouding by drugs or in any conversation. Now clearly, some of the intellectual processes such as counting or recording is extensional and is explicable by some computer model of the brain. This idea, although already foreshadowed by Descartes, is nevertheless a great achievement. It is very unsatisfactory all the same, as it excludes intensions. It explains pain as a signal but not as feeling. We know that these differ as we can separate them artificially. So how is pain explicable? What is required of its explanation to make it satisfactory? We do not know. Brain physiologists who look for an answer in the brain will fail to find it. For, they do not know what they are looking for; they do not know what the desire of an answer that will make it satisfactory.

Although we do not know the answer, we have an inkling of one: we hope to learn about the place of mind in nature. What minds add are feelings—feelings that remain when we ignore signals and pay attention to what they should say.

Meaning = semantics—syntax

We can see that when we observe the process in reverse, no matter how unlikely it is in fact. Consider seashells emitted by the waves at high tide that look like a written message; they are not messages, but we may use them as one! Less farfetched is the *objet trouvé*: it is an item that its finder declares an artwork. The addition that makes an item a message or an artwork is mental, intensional. Clearly, intensions makes them such, just as intensions renders some animals pets. This becomes obvious when we compare cats or dogs with collections of mushrooms or butterflies. What then is the addition to nature that intentions bring about? We have no clue.

Popper insisted that intensions are objective. He stressed that although we cannot imagine them without some souls hanging around, they are independent on any particular soul and so he called them a specific domain that he called World3. It includes abstract objects such as symphonies that differ from any of their space-time manifestations such as their written scores, performances or records. Popper placed the sensation of pain in world2, it seems, whereas the pain-signal belongs to world1. What signifies is that they are not reducible: no explanation of the feeling of pain will make it a part of physics.

A word about reduction. The most impressive reduction is of optics to electromagnetics. Optical phenomena are wavelike. It invites the theory that light comprises waves. These waves share the properties with electromagnetic waves. This invites the theory that light is an electromagnetic phenomenon. The theory becomes a part of physics when we consider Arclight as electromagnetic energy turned into light. Reduction is here complete explanation in the form, the whole of x is a case of y, where x and y are recognized fields of research. An effort—psychologism—to present all the human sciences as parts of psychology failed. Sociologism is a similar effort. It present all the human sciences as parts of sociology. It similarly failed. Another effort was the economism of Karl Marx that is a stringent version of sociologism, of course. Another effort is historicism that refers to political history and that Popper has tried to dismantle because he found it not only baseless but also immoral. Utilitarianism is an effort to reduce all ethics to the theory of conduct as based on the search for self-gratification. It led to much progress, and so it is morally commendable, even though as a reduction—as the idea that there is no morality other than self-gratification—it is morally objectionable and refuted (by self-sacrifice). We saw the same in the case of the reduction of the human sciences to computer science, of the reduction of biology (anthropology included) to computer science: as a partial reduction, it is terrific; as a reduction proper, it is a failure.

All aspects of humanity that belongs to physics—the weight and the pain-nerves of the human body—belong to World1; all aspects of psychology—the feeling of pain and even the variability of pain that depending on any social circumstances—belongs to World2; language belongs to World3. This is too neat: although language belongs to World3, speech belongs to World2 (think of speech therapy) and to World1 (think of speech as acoustics). Still, *grosso modo,* the picture seems unproblematic. What Popper intended to say is that it is impossible to reduce World2 to World1? Whence then did it come from?

This is a tricky question. Suppose we find a satisfactory answer to it. Whatever the explanation is, to be satisfactory, it should not include in it reference to World2 other than that World2 is x where x is a part of world1. This will be reduction, contrary to the claim that such a reduction is impossible.

What the previous short paragraph presents holds for all emergence: whatever is possible to explain as emergent is reducible to whatever allegedly gives rise to it and thus it will not be emergence. Hence, emergence with no reduction is impossible within the deductive-nomological theory of explanation. Hence, this theory does not allow for explanation of any emergence as emergence.

Another model of explanation is systemism (Mario Bunge). Qualities are systemic if and only if they belong to a set but not to its members. (Example: as the set of things is not a thing, being a thing is a systemic quality.) What set is a system is and remains unclear: it is not a clear-cut theory. (It is a heuristic, namely a research tool). Advocates of systemism consider a description explained once they have succeeded to integrate it into a system. This sounds too easy unless we offer some stringent specifications of what a system is and what integration is. Here, possibly, emergence can find its natural home. We must declare some sets real, since otherwise we will deny reality to anything other than what physics declares elementary particles. Consequently, you and I, as well as the pieces of furniture around us, will lose reality, as they are not elementary particles (but sets of them). Leading logician Willard Quine went further: he found it unavoidable to consider all sets real. Will this possibly apply to souls?

Can we say that souls are sets of sorts? Mario Bunge said that they are: souls are sets of brain characteristics. As long as these are not declared (reducible to) physical properties, Bunge is not a reductionist. This is not a solution to the mind-body problem but a scheme for it: find all the brain characteristics that comprise a soul. Popper had to agree; whether he would, we cannot know. The advantage of Bunge's systemism over Popper's institutionalism is in allowing systems to be not only institutions but also explicitly systems of institutions and people and whatever else they may involve. The advantage of Popper's institutionalism over Bunge's systemism is that it refers to institutions explicitly. Perhaps the theory of the one is implicitly also the theory of the other, so that there is no inherent difference between them, so that there is no reason to combine them. Nevertheless, obviously, it may be preferable to combine them explicitly, and there are quite a few default ways of doing so.

Bibliographical Note

Karl Popper presented and discussed his institutionalism in his *The Poverty of Historicism* of 1957 (first published in *Economica* in the early 40’s) and in his classics *The Open Society and Its Enemies* of 1945, particularly in Chapter 14, “The Autonomy of Sociology”

Mario Bunge presented his systemism in many publications of his, particularly in the part of his *Treatise on Basic Philosophy: Ontology II: A World of Systems* of 1979

See also Agassi, J., Jarvie, I.C. (eds.) *Rationality: the critical view*. Martinus Nijhoff, Dordrecht (1987)

Joseph Agassi M.Sc., Ph.D., Philosopher, Professor Emeritus Tel Aviv University, Israel, and York University, Toronto, Canada, *Membro Onorario del Corpo Accademico dell’Universita degli Studi “Gabriele d’Annunzio”*, Chieti-Pescara, Italia. Author or co-author of over twenty books, editor or co-editor of over ten books, author or co-author of over five hundred contributions to the learned press in diverse fields. https://www.tau.ac.il/~agass/index.html.

Objective Information, Intersubjectivity, and Popper's Three Worlds

Nir Fresco

1 Introduction

Information plays a key explanatory role in the sciences of mind and brain (Adams 2003; Artiga 2016; Cao 2012; Floridi 2014; Fodor 1990; Fresco et al. 2018; Fresco and Michael 2016; Mann 2018; Ramos 2014; Rathkopf 2017a; Ryder 2009; Scarantino and Piccinini 2010; Usher 2001). Objectivity is an ideal standard for scientific inquiry that justifies both the value of scientific knowledge and the confidence we place in such knowledge. But does this standard of objectivity force us to adopt only mind-*independent* notions of information in informational descriptions in the sciences of mind and brain? The answer advanced herein is negative. The motivation for the view that information is a "mind-independent commodity" seems clear: information can (supposedly) be received, stored, retrieved, passed around, and processed by minds/brains or sensorimotor systems. If "[i]nformation has to contribute to the origin of the mental" (Adams 2003, 495), then information had better existed independently of mental (or cognitive) states. A mind-dependent notion of 'information' can, nevertheless, be scientifically legitimate in informational descriptions (Fresco forthcoming). Epistemic objectivity suffices to ground the explanatory work that 'information' very often does in these sciences.

The present analysis does not specify *what* this information is. Rather, it advances the modest claim that *receiver-dependent* information can be objective in the Popperian sense. But which sense is that? Popper separately advances both World 3 and intersubjective objectivity. On a strict interpretation of World 3, objective information should be understood as information that is *totally autonomous* from Worlds 1 and 2. The contents of a book supposedly exists in World 3 even before anyone (in World 2) has read the book. On a more relaxed version of World 3, the objectivity of information should not be understood as 'total receiver-independence'. Such objectivity

---

N. Fresco (✉)
Ben-Gurion University of the Negev, Beer Sheva, Israel
e-mail: nfresco@bgu.ac.il

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8_19

can be understood as intersubjective agreement in the Kantian sense. “Plato wanted epistémé, true knowledge, to ensure that his views were objective, real, true. Kant gave up all hope for objectivity and settled once and for good for the intersubjective” (Agassi 1975, 4). Information, accordingly, can be scientifically objective so long as there is the potential of it being received and tested publicly. Nevertheless, this conclusion requires further modification to the standard intersubjectivity criterion.

The remainder of this chapter comprises four parts. Section 2 distinguishes between two senses of information that are typically used in the sciences of mind and brain. Section 3 advances the claim that the receiver-dependence of information need not undermine its scientific objectivity. The relation between Popperian intersubjectivity and World 3 objectivity is examined in Sect. 4. The chapter concludes—in Sect. 5—with the proposal to adopt a “slightly” modified version[1] of Popperian objectivity to ground the scientific objectivity of information in the sciences of mind and brain.

2 Which Information? Two Relevant Senses of ‘Information’

‘Information’ tends to be used in the sciences of mind and brain in two different, but related, senses. One sense is typically called ‘correlational’ or ‘natural’ information, referring to naturally co-occurring events. Examples include concentric rings in a tree and seasons of growth, smoke and fire, dark clouds and rain, and deer tracks in the snow and a deer. “For such events to carry information they must occur against a background of environmental regularity and consistency in which one event’s occurrence is [...] regularly and lawfully connected with another (as a kind)” (Adams 2010, 335–336). Dretske is famously known for requiring a “lawful correlation” between two events for one event to convey information about the other (1983). But this requirement is, arguably, too strong, and can be relaxed to simply specify that natural information increase the probability of the state of affairs it is about (Scarantino 2015). This type of information is often said to be factive: there cannot be correlational *mis*information (e.g., smoke cannot carry false information about there being fire nearby, even in the absence of fire) (Scarantino 2015, 430). The information that cognitive states “possess” often originates in correlational information (e.g., Adams 2010; Dretske 1981; Floridi 2014; Rathkopf 2017b; Shea 2018).

This brings us to the second sense of ‘information’, referring to the contents of sensorimotor, proprioceptive, and cognitive states. An organism may possess such information when sensory or proprioceptive input is received, “processed”, and used for inference or guiding behaviour. This is the result of such input having been *consolidated* into the sensorimotor, proprioceptive, or cognitive system, along with other priors, dispositions, beliefs, or knowledge; hereafter, we label it as ‘consolidated information’. How that input is integrated and subsequently made available for use by the receiver[2] is affected by specific background priors, dispositions, beliefs, or knowledge. Thus, consolidated information is different from 'natural information'. Some have chosen to call the former 'non-natural information' (Scarantino and Piccinini 2010; Lee 2019),[3] whereas others simply call it 'semantic information' (e.g., Adams 2010; Dretske 1981; Floridi 2011; Ramos 2014).[4]

It should be emphasised that consolidated information can be exploited even by sensorimotor organisms deprived of *cognitive* states. Such organisms may process consolidated information about their environments (an environmental feature, such as "no oxygen here") or the interaction of their bodies with their environments (e.g., switch the direction of flagellar motor rotation). However, why should we assume that a sunflower or an amoeba has any *semantic* or *cognitive* states that influence their behaviour? They can, nevertheless, be said to exploit the relevant consolidated information in accomplishing a specific task (e.g., the sunflower directing itself toward the sun, or the amoeba moving in the direction of food). The capacity to exploit consolidated information need not *in itself* imply the possession of cognition. (But this claim is not further examined in what follows).

For present purposes, we submit that only those physical events or patterns that are sensorily or perceptually available to a receiver can genuinely qualify *qua* information. The question is not whether the receiver potentially or actually exploits this input under the right conditions in making decisions about the state of the world and/or selecting which of the available behaviours are appropriate. Rather, the question is whether these events or patterns even qualify as *possible* inputs to the receiver to begin with. Elephants, for example, may be sensitive to nomic regularities transmitted via infrasound, whereas humans are not (without the necessary technological aids). A spatiotemporal correlation between smoke and fire cannot be perceived by an animal that can neither see nor smell. Thus, it cannot be informative *for* that animal.

But if information is indeed receiver-dependent, does it follow that information cannot be scientifically objective, because it is relativised? It should be clear, at this point, that *consolidated* information cannot be classified as a receiver-*independent* notion. Some philosophers argue that correlational information is a commodity that need not presuppose the existence of organisms or minds (e.g., Adams 2003; Dretske 1981; Floridi 2011; Millikan 2004). But since we are interested here in the scientific objectivity of information—despite it being receiver-dependent—let us simply assume that in the sciences of mind and brain *correlational* information, too, is receiver-dependent. (See Fresco (forthcoming) for an argument in support of this assumption). If both correlational and consolidated information are inherently relative to a receiver, and are at the core of informational explanations and models in the sciences of mind and brain, is the scientific objectivity of these explanations and models at risk?

3 Does Receiver-Dependent Information Undermine Scientific Objectivity?

Suppose that we accept that information *is* receiver-relative, does that by itself render information subjective, i.e., being based on the idiosyncratic beliefs, whims, biases, and/or preferences of a specific receiver? If that were the case, it would not exist beyond the receivers that interpret it (Pérez-Montoro 2007, 10). Should we then accept that "information is relative to the observer and for the very reason of the way it is created it is [necessarily] subjective" (Zaliwski 2011, 77)? The short answer is 'no'.

Explaining cognition and behaviour in informational terms requires dealing with receiver-dependent communication problems. For example, does a signal convey information for a receiver regardless of her prior informational state? The answer seems to be in the negative. (This chapter carries little information for non-English readers). If a single message is sent simultaneously from one source to two receivers with different informational states, can it be informative for one receiver but not the other? The answer seems to be in the affirmative. (If this chapter is sent electronically to a six-year old and to an expert on Popper's philosophy, it may be informative only to the latter). If a message is sent to the same receiver twice, is she equally informed upon receiving both messages? Again, the answer seems to be in the negative. (The first received message may be informative, whereas the second might only carry some meta-information, e.g., the transmitter still works). When information is understood functionally, such problems can, arguably, be addressed in a straightforward manner. For example, a single message can be informative for a receiver with the suitable "background knowledge" but not to an ill-equipped receiver.

Information—in a *teleosemantic* sense—can be said to exist relative to a receiver (e.g., some neural network, a sensorimotor system, a brain, or an entire organism) that has evolved to respond in a regular, causal manner to an event, environmental feature, or object by altering its internal state(s). Information understood *functionally* (see e.g., Cao 2012; Fresco et al. 2018; Jablonka 2002; Mann 2018; Rathkopf 2017b) presupposes the receiver's capacity to alter specific features of its internal state(s) as a function of systematic changes in the receiver's body or the surrounding environment. The receiver's being informed can be the result of phylogenetic or ontogenetic developmental processes. Thus understood, cues (e.g., deer tracks in the snow/deer) and signals (e.g., eagle-alarm-call/eagle) do not literally *contain* information as a mind-independent commodity, unlike mass or energy.

This teleosemantic approach to information, however, is by no means the received view in the literature. Dretske, for example, famously attempted to naturalise mental content by appealing to information as being mind-*independent*. Nevertheless, this unsuccessful attempt provides a glimpse into the difficulty of explaining cognition and behaviour by appealing to mind-independent information. "[W]hat information is transmitted [by a signal] may depend on what the receiver *already knows* about the possibilities existing at the source" (Dretske 1981, 65, italics added). Dretske argued that this relativisation *does not* undermine the mind-independence of information:

the flow of information occurs even "without conscious agents who know things" (1983, 57). But *if* this prior knowledge can indeed be explained by appealing to some neural circuitry or non-intentional states, his definition of information should have been reformulated without relying on the receiver's prior knowledge.

The upshot of this brief analysis is that a receiver—properly constrained, of course—can be studied as an object of scientific inquiry as any other physical system—this is indeed a basic assumption of the cognitive sciences. The receiver's conditional responses to the environment can be studied just as any causal system whose functioning depends on background conditions, and so on. There seems to be little reason to treat the scientific objectivity of organism-environment interaction—when analysed in informational terms—any differently than other epistemic problems in science. Scientific models are regularly evaluated for their degree of fit with data. Insofar as truth-evaluable predictions about some target phenomenon can be made on the basis of sample data, then the model is typically deemed epistemically objective. Understanding 'information' as a scientifically acceptable, receiver-dependent notion lays the path for alternative naturalistic frameworks for explaining cognition.

4 Popperian Intersubjectivity and World 3 Objectivity

To defend the claim that information can indeed be both receiver-dependent and scientifically objective, we appeal to the Popperian notion of objectivity. In his 1934 book, "*The Logic Of Scientific Discovery*", Popper, it may be argued, "provided the first philosophically tenable account of information transfer or, as he then called it, of the acquisition and growth of knowledge" (Munz 1993, 143). As is well known, he defended the falsifiability, instead of the verifiability, principle as the demarcation criterion of science,[5] but, importantly, as the principle underlying information transfer, too. Popper argued that scientific discovery has to exceed the *subjective* features of discovery. These subjective features are to be eliminated through testing procedures (rigorous attempts at falsification) for information to be objective. Such testing procedures play an important role in both intersubjective agreement and World 3—to which we turn next.

Let us first examine Popper's Three Worlds view in relation to the objectivity/subjectivity of information. On this view, there exists a hierarchy of worlds: a world of physical phenomena (*World 1*), a world of behavioural and mental phenomena, including thoughts and beliefs (*World 2*), and a world of objective rational knowledge, which includes theories, propositions, arguments and problems (*World 3*). World 1 makes possible behaviour, cognition and consciousness in World 2. The autonomous status of World 3 and its putative casual relations to the other two worlds have been criticised on various grounds. For one thing, problems and other abstract entities in World 3 exist prior to the subject's grasp of them, and are, thus, merely discoverable (Parusnikova 1990, 266).

For another, if World 3 objects are abstract, how can they causally influence physical events in Worlds 1 and 2? Consider, first, informational objects that may belong to two worlds simultaneously. Journals, books, and libraries exist both in World 1 and World 3. Insofar as they are *physical* objects they belong to the former: "they are subject to the physical restrictions or physical laws of world 1" (Popper 1973, 20). Yet, they also belong to the latter by being "subject to the restrictions and the valuations of world 3, such as the laws of logical consistency and the value of having *informative content*" (ibid., italics added). Whilst two printed copies of the same book are *different* as World 1 objects, if their *contents* are the same, then they are *identical* as objects in World 3. "[W]orld 3 objects can have a strong causal influence upon world 2 processes. And if a newly discovered world 3 problem […] is published, then the causal influence extends even into world 1" (Popper 1973, 23). To take another example, the planning and construction of skyscrapers requires first the understanding of a World 3 theory, and then an interaction between World 2 planning, and the internal restrictions of Worlds 1 and 3 (Popper 2012, 117).

Popper seems to endorse an *active* Darwinian view with respect to the growth of knowledge. "Animals […] are problem-solvers. And they solve their problems by the method of competitive tentative solutions and the elimination of error" (Popper 1972, 145). But do animals indeed partake in the three worlds dynamics irrespective of humans that describe their behaviour? Popper seems to conceive of "world 3 as a *human* product. Critical *human* thinking (world 2) solves problems and hence produces new ideas (world 3)" (Parusnikova 1990, 266, italics added). And if "[a]ny problem entails conscious formulation, intellectual reflection, and hence its existence cannot be independent of man" (Parusnikova 1990, 267), then it would seem that the answer is negative. However, nonhuman animals solve problems all the time *regardless* of whether we observe them in doing so. For example, Betty—a famous New Caledonian crow—bent a straight piece of wire into a hook in order to retrieve a piece of food that was trapped in a plastic tube. (To be sure, Betty *was* observed by human scientists who then reported these findings). Whether or not Betty had *a hook in mind* when preparing the tool, it used an unorthodox solution to solve a novel problem (Kacelnik et al. 2009).

What about Popper's emphasis on World 3 objects being (abstract) instances of *symbolic* information, such as statements, propositions, and arguments? World 2 mental objects gain their objectivity through their formulation in human language. "[A] thought once it is formulated in language, becomes an *object* outside ourselves" (Popper 2012, 118). If all World 3 objects are symbolic, then it seems that the applicability of World 3 objectivity to information in the sciences of mind and brain would be limited at best (i.e., to a subset of human consolidated information). Betty the crow certainly does not convey her solution to the food-extraction problem in any symbolic manner. Yet, whether or not this specific solution *has been* socially learned, "[i]t is well established that the behaviour of a single individual can spread and become established in a population" (Kacelnik et al. 2009, 525). And Popper seems to acknowledge that theory-like expectations occur even at a purely *perceptual* level: "sense organs […] incorporate […] theory-like expectations. [They …] are prepared to react to certain selected environmental events–to those events which they ‘expect’” (Popper 1972, 145).

In principle, then, Popper’s theory of information transfer may be deemed “equally applicable to animal and human learning […, thereby …] abolish[ing] the gap which the conventional wisdom of philosophers had insisted exists between prehuman and human information gathering” (Munz 1993, 144–145). Whether or not Popper’s World 3 was *originally* intended to also include nonhuman animal problems (presumably not) is less important for present purposes and addressing this question exceeds the scope of this chapter. However, insofar as a properly modified World 3 can accommodate non-symbolic information and nonhuman animal problems, it can arguably be used to ground the objectivity of receiver-dependent information. Additionally, the dynamics between the three worlds would also require some kind of a “feedback principle” (Parusnikova 1990, 268): a proper “elaboration of the mutual interaction of the subjective and the objective governing the process” (Parusnikova 1990, 266).[6]

Where does that leave us? Objective knowledge, according to Popper, should be set apart from subjective knowledge that belongs to World 2. “Knowledge in this objective sense is totally independent of anybody’s claim to know; also it is independent of anybody’s belief, or disposition to assent […] Knowledge in the objective sense is *knowledge without a knower*” (Popper 1972, 109, italics original). Once a theory or problem becomes publicly available—by being communicated–it gains its autonomy. “When entertained by some mind, the statements we consider may be prey to any number of subjective embellishments. Nonetheless as statements they are objective public items and they belong to no one in particular” (Miller 2006, 103). In that limited sense, Popper’s view resonates in Frege’s: “I understand by a thought not the subjective performance of thinking but its objective content, which is capable of being the common property of several thinkers” (Frege 1892, 62).

This seems problematic, however, since the Popperian view that there can be knowledge without a knower entails that there can be information without an informee (i.e., a receiver). Which sense of objectivity can we endorse, then, for *receiver-dependent* information? Suppose that “[a]ll our machines and tools are destroyed [… along with] all our subjective learning […] But *libraries and our capacity to learn from them* survive […] after much suffering, our world may get going again” (Popper 1972, 108). In this thought experiment, “objective” knowledge survives in libraries and allows civilisations to be rebuilt by consumers of knowledge. The library is akin to an external memory; the *information* in it can be read by anyone with the right perceptual apparatus and language skills (Rowbottom 2014). Now suppose—beyond what Popper suggested—that all libraries survive, and yet they are *never* found again by any civilised successors of ours or other potential interpreters (e.g., higher apes). If there are neither *actual* nor *potential* receivers of information in the library, what kind of information is it? Likewise, Popper would likely have said that the library ceases to be a *library* (Agassi, personal communication).

There needs to be at least the potential for the interpretation of information for something to qualify as information to begin with. Pebbles arranged on the beach to form the word S.O.S—whether by chance or not—cannot convey any information in the absence of receivers; they are just pebbles. Information is produced by humans and nonhuman animals interacting with their environment, rather than through some organism-independent process. And, indeed, signalling, on Popper's view, is a lower function common to both human and nonhuman animal communication. "The signalling […] function is […] obvious: we do not call any symptom linguistic *unless* we assume that it *can release a response in another organism*" (Popper 1972, 120, italics added). Popper also conceded that "in order to belong to the third world of objective knowledge, a book should–in principle […]–be capable of being grasped [or interpreted]" (1972, 116). But his "attempt to desubjectivize the realm of objective knowledge leads to a philosophically unbalanced position" (Parusniková 2016, 313).

Popperian objectivity can be relaxed to be understood simply in terms of *intersubjective agreement*, thereby requiring at least two receivers. Whilst, according to logical empiricism, every evidence cohering with a given theory may corroborate that theory, in Popper's view, a theory is corroborated by having withstood severe falsification tests (1965, chap. 10). *Objective* knowledge is that which becomes publicly available and agreeable upon by the relevant community. Similarly, information is scientifically objective when it is not specific to a single receiver in World 2, but sharable amongst many and is open to testing (e.g., by observation, and experimentation). This holds true whether or not the receivers believe the information to be veridical. And, certainly, World 3 also includes false theories, and even inconsistent arguments and theories (Popper 2012, 115).

It should be noted, however, that Popper never explicitly discussed the criteria for belonging to World 3 (Agassi, personal communication). He clearly rejected identifying World 3 with the *Platonic heaven* (Popper and Eccles 1977, 43–44) and with *Frege's third realm*, insofar as it too is conventionally platonic, as "thoughts are timeless entities, unaffected in their essences by the activities of human subjects" (Currie 1989, 419). It is not clear whether the intersubjectivity criterion was intended to be one of the admission criteria for belonging to World 3. For, to reprise, unlike the Platonic realm—World 3 also includes *false* conjectures. Even basic facts and conjectures reside in World 3 and not just potential and actual theories. Belonging to "World 3 means nothing less and nothing more than an 'objectified thought,' an idea or a hypothesis that has a fixed objectified form—oral or preferably written and printed" (Parusniková 2016, 309).

Popperian intersubjectivity serves as a means for resisting the psychologism of knowledge. As such, it targets assertions, theories, arguments, and other *linguistic* constructs. Once a thought "is formulated in language, [it] becomes an *object* outside ourselves. Such an object can then be inter-subjectively *criticized*—criticized by others as well as by ourselves" (Popper 2012, 118, italics original). And with such intersubjective criticism that "emerges with human language [… also] emerges the human World 3, the world of objective standards and of the contents of our subjective thought processes" (ibid.). Nevertheless, just as a theory can be tested and refuted, so can solutions to problems in general, as well as simple and complex expectations in humans and nonhuman animals.

In that respect, intersubjective testing, too, can be expanded (as suggested above for World 3) to also apply to information that is not merely human-based consolidated information and to nonhuman animal-based information. Does that idea cohere with the Popperian view of objective knowledge? Arguably, yes. For Popper claims that "[s]cientists, like all organisms, work with the method of trial and error. The trial is the *solution to a problem*" (1999, 38, italics original). Both a low-level organism (e.g., an amoeba) and a human—and every organism in between—are problem solvers. Problem solving—including a scientific theory as the penultimate, tentative product of this process—is performed by means of conjectures and refutations. For the amoeba, to be sure, this possibly means no more than attaining an immediate goal in a given problem-situation.

Whilst the "correction" of a false *scientific* hypothesis or theory results in the eradication of the hypothesis (or theory), that correction "usually means eradication of the *organism*" (ibid., italics added). What typically "corresponds to so-called knowledge, to conjecture or hypothesis [… is] *expectation* [… or …] a state of the organism in which it prepares for a change (or no change) in its surroundings" (ibid., italics original). A scientist looks for mistakes, since "his theory is not part of himself but an object that he can consciously investigate and [intersubjectivley] criticize" (Popper 1999, 39). A primitive organism, such as an amoeba, on the other hand, "is eliminated when it *makes* mistakes. If it is conscious it will be afraid of mistakes" (ibid., italics original).

The upshot of the present analysis is that objectivity-as-intersubjectivity arguably applies to any problem situation and can be used as a criterion for grounding the scientific objectivity of receiver-based information. This criterion of objectivity-as-intersubjectivity applies to low-level organisms solving immediate problems, on one extreme, and to scientists formulating new theories, on the other extreme. Nevertheless, one should concede that "[t]he subjective is inherent in the objective because [the former] produces [the latter] and the objective is inherent in the subjective because it determines its focus" (Parusniková 2016, 313). The objectification of information (also) resides in cognitive and other problem-solving capacities that are the inhabitants of World 2. The subjective being inherent in the objective (World 3) need not contradict the objectified status of World 3 (Parusniková, personal communication).

5 Applying Popperian Objectivity to Receiver-Based Information

Let us now observe how Popperian objectivity can be applied to receiver-based information that is, arguably, the basis for informational explanations and models in the sciences of mind and brain. How can intersubjective testing properly expanded cater for both correlational and consolidated information? Recall, first, that, on the view advanced here, even correlational information is assumed to be receiver-dependent, since such information is not simply a natural relation between physical events. It may be objective, and yet "relativized and mind-dependent" (Scarantino 2015, 438)—just as a tree is *objectively climbable* relative to a specific vervet monkey.

Second, to reprise, the processing of consolidated information may also occur in nonhuman animals, and thus should not be equated with symbolic information.

The most general way to see how Popperian objectivity can be applied to these types of information is to consider human and nonhuman animal cognition and behaviour[7] as *problem solving*. In solving problems—even when simply watching a movie (where both perceptual and cognitive processing occurs), or when bacteria move towards nutrients and away from toxins (where perceptual processing suffices)—information transfer occurs. And when this (receiver-based) information is transferred or processed, it may be objective (despite being receiver dependent) in being publicly available and (intersubjectively) testable.

Let us consider a few "information processing" examples. Understood functionally, a dark cloud is only informative for an ape that has learned that rain is likely to follow (the dark cloud is a conditioned stimulus). That is, it is not *only* in virtue of a probabilistic regularity between the two abiotic events that the dark clouds are informative for the receiver. A baby ape that has yet to learn this regularity will not infer anything about the coming rain when observing dark clouds. A mature ape, however, will not only see the dark clouds in the horizon, but it will also, if needed, take appropriate action (e.g., seek shelter from the rain) on the basis of the predictive association formed between the clouds and rain. A prelinguistic baby crying to his parent "expects to be cared for and nursed, and soon to be smiled at" (Popper 1999, 39). Even before the baby understands language, he exchanges information with his parents. Receiver-dependent information in both of these cases is objective insofar as it is not specific to any one receiver, and is sharable amongst many receivers (whether or not it is believed to be *true*).

Consider, next, another example of conspecific communication in the animal kingdom. Suppose that one vervet (V_1) produces a leopard alarm call (c) and another vervet (V_2) responds by climbing up the nearest tree (Fresco forthcoming). Let us also grant that the information in c is relative to V_1 and V_2 (whilst the presence of the leopard underpins the probabilistic regularity in nature). If there were no fact of the matter about V_1 sending information to V_2 about an external referent (i.e., the leopard approaching), why do such calls typically elicit distinct sets of vervet behavioural responses? Two observing scientists should agree (at least eventually in the light of evidence and following rational criticism) about the information communicated between V_1 and V_2. That is, assuming a complete naturalistic specification is available of the states of V_1 (e.g., spotting the leopard) and V_2 (e.g., its attention drawn to the call), the auditory properties of c (e.g., the length and amplitude of the alarm call), and, of course, the leopard's approach. The information conveyed between V_1 and V_2 is receiver-dependent: even though there is an *innate* basis for the production and usage of alarm calls in vervets, there is much fine-tuning through individual and social learning in producing and responding to these calls (Seyfarth and Cheney 2010).

The proposed means to objectify information, drawing on the aforementioned analysis of Popperian objectivity, is to make the information concerned visible as a collective enterprise—where assessing the competence and character of the participants is crucial. In science, the objectification of information is a deliberative process, whereas in the animal kingdom it is the result of selectional pressure. The bias or idiosyncrasy of judgments and opinions of any particular individual are put in check by producing and validating information through public discourse (or again: selectional pressure) and by following certain epistemic norms, which are institutionally embedded and reinforced.[8] As long as the zoologists' reports about the vervets are interconsistent, there can be intersubjective agreement between them about the external conditions (e.g., the behaviours of V_1 and V_2, the leopard approaching, the alarm call, and so on). Whilst the leopard alarm call indeed has a subjective reference—in V_1 producing the call and V_2 receiving it—that is private (i.e., it exists in World 2), it also has an intersubjective reference to the physical conditions of the relevant events that other observers (both zoologists and vervets) can affirm and are, therefore, confirmable by means of intersubjective agreement (Freeman 1973, 169). Of course, the zoologists' intersubjective testing—by means of subjecting their hypothesis to the test—differs from that of the vervets—by means of phylogenetic and ontogenetic selection; vervets failing to respond well to leopard alarm calls will perish.[9]

Thus understood, the relevant sense of objectivity of information cannot imply a *total* independence of any receivers (i.e., knowledge without a knower). At the very least, the potentiality of receivers who can interpret information is paramount. Absent that potentiality, as in the second version of the library thought experiment discussed above, there can be no information in libraries, books, or digital databases. The objectification of information through intersubjective testing unfolds as a dynamic interaction between Worlds 1, 2, and 3. Such view is consistent with a functional notion of information as a triadic relation amongst a receiver, a signal (e.g., vervet alarm call) or cue (e.g., smoke or deer tracks in the snow), and some object, feature, or state of affairs (e.g., the putative presence of leopard, fire, or deer, respectively).

The view advanced here differs, for example, from Mingers' intersubjective account of information-generated *meaning* (1995). In his view, causal events convey mind-independent information (i.e., correlational information). However, it is only *meaning*—generated *from* correlational information—that can ever be intersubjective. Humans, in Mingers' view, cannot access mind-independent information, simply because we are necessarily embedded within a domain of meaning. (Bar-Am calls this 'absolute information'—a basic aspect of reality-as-it-is-in-itself, and the *potential* for content (2016, 93)). The same mind-independent information, so Mingers claims, triggers various meanings in different receivers: some meanings may be idiosyncratic but others intersubjective. It is not clear, however, what such *meaning* amounts to, but certainly consolidated information need not assume it, and receivers need not be confined to humans.

The proposed view of objectivity need not fall prey to behaviourism either. Behaviourists would argue that (a) only overt behaviour can be intersubjectively agreed upon, and (b) there is no need to go beyond overt behaviour. Functionalists would counter-argue, however, that behaviours are also the result of complex internal states and processes. The underlying assumption is that information is scientifically objective when "under the specified conditions it will be replicated by any normal subject every time the prescribed conditions are present regardless of what the subject consciously wills about it" (Freeman 1973, 170). Receiver-based information is not confined to *actually observed* phenomena. It can causally contribute to the *biological function* of the receiver.

Still, an obvious challenge to intersubjective agreement as a marker of scientific objectivity is that it is supposedly too weak as a criterion of objectivity. For it "cannot guarantee that one has gotten at something real; there is always the chance of a group illusion" (Douglas 2004, 463). A group of people might—repeatedly and systematically—agree on the basis of brain-washing or the reading of tea leaves that plausibly have little, if anything, to do with the truth (Rowbottom 2008, 128). Illusory and hallucinatory perceptions can still be treated as non-veridical information that is supposedly scientifically objective under the right conditions: there can in principle be intersubjective agreement amongst normal subjects about illusory perceptions. (Just consider the visual illusion of a pencil bending in a glass of water).

The short reply is that an appeal to intersubjective testing can reveal that some subjects make use of non-veridical/erroneous or partial information (Fresco forthcoming). A supplementary, yet veridical, report that is based on additional information (perhaps, such that it is not readily available to the subjects of illusory percepts) can establish the non-veridicality of the illusory perceptions. The experienced illusion can still be classified as objective, insofar as any normal subject experiences it similarly under the same conditions—it is a shareable event (Freeman 1973, 172–173). Intersubjective criticism may result in the production of ideas that would not otherwise have arisen, such as finding new ways to question the accuracy of some perceived phenomenon, or that specific theories may account for the phenomenon concerned. Moreover, error correction, in principle, is more successful in larger groups: intersubjective agreement may be conditional, for example, on ironing out any identified mistakes (Rowbottom 2008, 129).

In sum, objectivity is undoubtedly the goal of all science, and, receiver- or mind-dependent information—appealed to in the sciences of mind and brain—should be likewise objective. The scientific objectivity of such information can be attained through intersubjective testing, or, when "actualised" information in World 3 stands in the right relation to World 1 (and verifying that relation requires a dynamic process that unfolds in World 2). Popper provided us with a promising route to ensure the objectivity of functional information despite its inevitable relativity; this route, however, needs some adjusting: there can be no (functional) information without an informee.

Notes

1. How slight that modification is will be left to the reader to decide.
2. 'Receiver' and 'receiving-organism' are hereafter treated synonymously. For a useful comparison of the two types of receivers (e.g., a metal detector and a football player) of semantic information born by representation see Colombo (2010).
3. This notion tracks the Gricean distinction between natural and non-natural meaning (Grice 1957). In contrast to natural information, there *can be* non-natural misinformation.

4. Both labels are problematic (see Cao 2012; Rathkopf 2017b for a critique of the “semantic” label).
5. Whilst the falsifiability criterion, thereby, solves some age-old problems in philosophy, it also raises a number of new problems. We will not discuss them here.
6. Some have argued, however, that it is not clear whether the elaboration of Popper’s three-worlds view really leaves room for a critical cognitive subject. “His requirement of the knowing subject is not innerly united with his objectivist approach” (Parusnikova 1990, 267). Supposedly, even the more relaxed characterisation that Popper advances in (Popper and Eccles 1977) just introduces “more problems and requires a fundamental revision of the initial conception of World 3. The initially postulated autonomy of the third world becomes doubly untenable when social institutions are involved” (Parusniková 2016, 314).
7. This characterisation is intended to be broad enough so as to include even the simple behaviour of amoebas.
8. See Popper and Eccles (1977), for an analysis of the scientist’s task of formulating a theory as an interaction between Worlds 1, 2 and 3.
9. See, Fresco (forthcoming) for an analysis of another route of objectivty of information in the context of nonhuman animals in terms of response dependence.

References

Adams, F.: The informational turn in philosophy. Mind. Mach. **13**, 471–501 (2003). https://doi.org/10.1023/A:1026244616112

Adams, F.: Information and knowledge à la Floridi. Metaphilosophy **41**, 331–344 (2010). https://doi.org/10.1111/j.1467-9973.2010.01630.x

Agassi, J.: Subjectivism: from infantile disease to chronic illness. Synthese **30**, 3–14 (1975). https://doi.org/10.1007/BF00485289

Artiga, M.: Teleosemantic modeling of cognitive representations. Biol. Philos. **31**, 483–505 (2016). https://doi.org/10.1007/s10539-016-9525-3

Bar-Am, N.: In search of a simple introduction to communication. Springer (2016)

Cao, R.: A teleosemantic approach to information in the brain. Biol. Philos. **27**, 49–71 (2012). https://doi.org/10.1007/s10539-011-9292-0

Colombo, M.: How “authentic intentionality” can be enabled: a neurocomputational hypothesis. Mind. Mach. **20**, 183–202 (2010). https://doi.org/10.1007/s11023-010-9192-0

Currie, G.: Frege and Popper: two critics of psychologism. In: Gavroglu, K., Goudaroulis, Y., Nicolacopoulos, P. (eds.) Imre Lakatos and Theories of Scientific Change, pp. 413–430. Springer, Netherlands, Dordrecht (1989)

Douglas, H.: The irreducible complexity of objectivity. Synthese **138**, 453–473 (2004). https://doi.org/10.1023/B:SYNT.0000016451.18182.91

Dretske, F.I.: Knowledge & the flow of information. The MIT Press, Cambridge (1981)

Dretske, F.I.: Précis of knowledge and the flow of information. Behav. Brain Sci. **6**, 55 (1983). https://doi.org/10.1017/S0140525X00014631

Floridi, L.: The philosophy of information. Oxford University Press, Oxford (2011)

Floridi, L.: Perception and testimony as data providers. Log. & Analyse. **57**, 151–179 (2014)

Fodor, J.A.: Information and representation. In: Hanson, P.P. (ed.) Information, Language, and Cognition. University of British Columbia Press, Vancouver (1990)

Freeman, E.: Objectivity as "intersubjective agreement": Monist. **57**, 168–175 (1973). https://doi.org/10.5840/monist19735722

Frege, G.: On sense and reference. In: Geach, P.T., Black, M. (eds.) Translations from the Philosophical Writings of Gottlob Frege. Blackwell, Oxford, England (1892)

Fresco, N.: Information, cognition and objectivity. Am. Philos. Q. (forthcoming)

Fresco, N., Ginsburg, S., Jablonka, E.: Functional information: a graded taxonomy of difference makers. Rev. Philos. Psychology. (2018). https://doi.org/10.1007/s13164-018-0410-7

Fresco, N., Michael, M.: Information and veridicality: information processing and the bar-Hillel/Carnap paradox. Philos. Sci. **83**, 131–151 (2016). https://doi.org/10.1086/684165

Grice, H.P.: Meaning. Philos. Rev. **66**, 377–388 (1957). https://doi.org/10.2307/2182440

Jablonka, E.: Information: its interpretation, its inheritance, and its sharing. Philos. Sci. **69**, 578–605 (2002). https://doi.org/10.1086/344621

Kacelnik, A., Chappell, J., Kenward, B., Weir, A.A.S.: Cognitive adaptations for tool-related behavior in new caledonian crows. In: Wasserman, E.A., Zentall, T.R. (eds.) Comparative Cognition Experimental Explorations of Animal Intelligence, pp. 515–528. Oxford University Press (2009)

Lee, J.: Structural representation and the two problems of content. Mind Lang. **34**, 606–626 (2019). https://doi.org/10.1111/mila.12224

Mann, S.F.: Consequences of a functional account of information. Rev. Philos. Psychology. (2018). https://doi.org/10.1007/s13164-018-0413-4

Miller, D.: Out of error: further essays on critical rationalism. Ashgate Publishing, Aldershot, UK (2006)

Millikan, R. G.: Varieties of Meaning: The 2002 Jean Nicod Lectures. The MIT Press, Cambridge, MA (2004)

Mingers, J.C.: Information and meaning: foundations for an intersubjective account. Inf. Syst. J. **5**, 285–306 (1995). https://doi.org/10.1111/j.1365-2575.1995.tb00100.x

Munz, P.: Philosophical Darwinism: on the origin of knowledge by means of natural selection. Routledge, New York (1993)

Parusnikova, Z.: Popper's world 3 & human creativity. Int. Stud. Philos. Sci. **4**, 263–269 (1990). https://doi.org/10.1080/02698599008573366

Parusniková, Z.: The devaluation of the subject in Popper's theory of world 3. Philos. Soc. Sci. **46**, 304–317 (2016). https://doi.org/10.1177/0048393116639227

Pérez-Montoro, M.: The phenomenon of information: a conceptual approach to information flow. The Scarecrow Press, Lanham, MD (2007)

Popper, K.R.: Conjectures and refutations: the growth of scientific knowledge. Harper & Row, New York (1965)

Popper, K.R.: Objective knowledge: an evolutionary approach. Oxford University Press, Oxford (1972)

Popper, K.R.: Indeterminism is not enough. Encounter. 20–25 (1973)

Popper, K.R.: All life is problem solving. Routledge, London (1999)

Popper, K.R.: The open universe: an argument for indeterminism from the postscript to the logic of scientific discovery. Routledge (2012)

Popper, K.R., Eccles, J.C.: The self and its Brain. Springer, Berlin, Berlin, Heidelberg (1977)

Ramos, R.T.: The concepts of representation and information in explanatory theories of human behavior. Front. Psychol. **5**, (2014). https://doi.org/10.3389/fpsyg.2014.01034

Rathkopf, C.: Neural information and the problem of objectivity. Biol. & Philos. **32**, 321–336 (2017a). https://doi.org/10.1007/s10539-017-9561-7

Rathkopf, C.: What kind of information is brain information? Topoi. (2017b). https://doi.org/10.1007/s11245-017-9512-6

Rowbottom, D.P.: Intersubjective corroboration. Stud. Hist. Philos. Sci. Part A **39**, 124–132 (2008). https://doi.org/10.1016/j.shpsa.2007.11.010

Rowbottom, D.P.: Information versus knowledge in confirmation theory. Log. & Anal. **57**, 137–149 (2014)

Ryder, D.: Problems of representation I: nature and role. In: Symons, J., Calvo, P. (eds.) The Routledge Companion to Philosophy of Psychology, pp. 233–250. Routledge, London (2009)

Scarantino, A.: Information as a probabilistic difference maker. Australas. J. Philos. **93**, 419–443 (2015). https://doi.org/10.1080/00048402.2014.993665

Scarantino, A., Piccinini, G.: Information without truth. Metaphilosophy **41**, 313–330 (2010). https://doi.org/10.1111/j.1467-9973.2010.01632.x

Seyfarth, R.M., Cheney, D.L.: Production, usage, and comprehension in animal vocalizations. Brain Lang. **115**, 92–100 (2010). https://doi.org/10.1016/j.bandl.2009.10.003

Shea, N.: Representation in cognitive science. Oxford University Press, New York (2018)

Usher, M.: A statistical referential theory of content: using information theory to account for misrepresentation. Mind Lang. **16**, 311–334 (2001). https://doi.org/10.1111/1468-0017.00172

Zaliwski, A.S.: Information—is it subjective or objective? tripleC: communication. Cap.Ism & Crit. **9**, 77–92 (2011)

Nir Fresco is a senior lecturer at the Department of Brain & Cognitive Sciences, Ben-Gurion University of the Negev, Israel. One branch of his research focuses on understanding—from informational and computational perspectives—how information processing contributes to our cognitive life. Another branch of research combines philosophical and empirical methods to study the relation between automaticity, habit formation, and skill acquisition in humans. E-mail: nfresco@bgu.ac.il

Index

Z. Parusniková and D. Merritt (eds.), *Karl Popper's Science and Philosophy*,
https://doi.org/10.1007/978-3-030-67036-8