H. Atmanspacher · H. Primas (Eds.)

Recasting Reality

Wolfgang Pauli's Philosophical Ideas and Contemporary Science

Springer

Recasting Reality

Harald Atmanspacher · Hans Primas (Eds.)

Recasting Reality

Wolfgang Pauli's Philosophical Ideas and Contemporary Science

Springer

Dr. Harald Atmanspacher
Institut für Grenzgebiete der
Psychologie und Psychohygiene e. V.
Abt. Theorie und Datenanalyse
Wilhelmstr. 3a
79098 Freiburg
Germany
haa@igpp.de

Prof. Hans Primas
Kusenstr. 21
8700 Kuesnacht
Switzerland

ISBN: 978-3-540-85197-4 e-ISBN: 978-3-540-85198-1

Library of Congress Control Number: 2008934463

Cover design: deblik, Berlin

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Contents

List of Contributors

Harald Atmanspacher
Institute for Frontier Areas
of Psychology
D–79098 Freiburg
Germany
`haa@igpp.de`

Giuliana Galli Carminati
Department of Psychiatry
University Hospitals of Geneva
CH–1225 Geneva
Switzerland
`Giuliana.GalliCarminati@hcuge.ch`

Thomas Filk
Parmenides Center
for the Study of Thinking
D–80333 Munich
Germany
`Thomas.Filk@t-online.de`

Georg Franck
Department of Computer Aided
Architecture and Planning
Vienna University of Technology
A–1040 Vienna
Austria
`franck@iemar.tuwien.ac.at`

Domenico Giulini
Max-Planck-Institute
for Gravitational Physics
D-14476 Golm/Potsdam
Germany
`domenico.giulini@aei.mpg.de`

Peter beim Graben
School of Psychology
and Clinical Language Sciences
University of Reading
Reading RG6 6AH
United Kingdom
`p.r.beimgraben@reading.ac.uk`

Joachim Klose
Konrad-Adenauer Stiftung
D–01097 Dresden
Germany
`Joachim.Klose@kas.de`

Günther Knoblich
School of Psychology
University of Birmingham
Birmingham, B15 2TT
United Kingdom
G.Knoblich@bham.ac.uk

François Martin
Laboratoire de Physique Théorique
et Hautes Energies
Universités Paris 6 et 7
F–75252 Paris Cedex 05
France
`martin@lpthe.jussieu.fr`

Karl von Meyenn
Institute for Theoretical Physics
University of Ulm
D–89081 Ulm
Germany

Arthur I. Miller
Department of Science
and Technology Studies
University College
London NW3 2NR
United Kingdom
`a.miller@ucl.ac.uk`

Rafael Núñez
Department of Cognitive Science
University of California at San Diego
La Jolla, CA 92093–0515
USA
`nunez@cogsci.ucsd.edu`

Michael Öllinger
Parmenides Center
for the Study of Thinking
D–80333 Munich
Germany
`michael.oellinger`
`@parmenides-foundation.org`

Hans Primas
ETH Zurich,
CH–8092 Zurich
Switzerland
`primas@phys.chem.ethz.ch`

Hartmann Römer
Institute of Physics
University of Freiburg
D–79104 Freiburg
`hartmann.roemer`
`@physik.uni-freiburg.de`

William Seager
Department of Philosophy
University of Toronto
Ontario
Canada M1C 14A
`seager@utsc.utoronto.ca`

Linda van Speybroeck
Department of Philosophy
and Moral Science
University of Ghent
FWO–Flanders
B–9000 Ghent
Belgium
`linda.vanspeybroeck@UGent.be`

Max Velmans
Department of Psychology
Goldsmiths College
University of London
London SE14 6NW
United Kingdom
`psa01mv@gold.ac.uk`

Introduction

Harald Atmanspacher[1] and Hans Primas[2]

[1] Institute for Frontier Areas of Psychology, Freiburg, Germany, `haa@igpp.de`
[2] ETH Zurich, Switzerland, `primas@phys.chem.ethz.ch`

The notion of reality is of supreme significance for our understanding of nature, the world around us, and ourselves. As the history of philosophy shows, it has been under permanent discussion at all times. Traditional discourse about reality covers the full range from basic metaphysical foundations to operational approaches concerning human kinds of gathering and utilizing knowledge, broadly speaking epistemic approaches. However, no period in time has experienced a number of moves changing and, particularly, restraining traditional concepts of reality that is comparable to the 20th century.

Early in the 20th century, quite an influential move of such a kind was due to the so-called Copenhagen interpretation of quantum mechanics, laid out essentially by Bohr, Heisenberg, and Pauli in the mid 1920s. Bohr's *dictum*, quoted by Petersen (1963, p. 12), was that "it is wrong to think that the task of physics is to find out how nature is. Physics concerns what we can say about nature." Although this standpoint was not left unopposed – Einstein, Schrödinger, and others were convinced that it *is* the task of science to find out about nature itself – epistemic, operational attitudes have set the fashion for many discussions in the philosophy of physics (and of science in general) until today.

Moreover, epistemically dominated directions have taken over in other disciplines as well. The *linguistic turn*, often ascribed to the influence of Wittgenstein in the 1930s and 1940s, is of key significance in this context. It was first spelled out explicitly by Rorty (1967) in his anthology "The Linguistic Turn: Essays in Philosophical Method". It demands, similarly to Bohr's appeal, to give up on asking how the world is but, rather, concentrating on how it is described. Philosophy of language becomes a central field in analytic philosophy, generating vast influences on phenomenology, anthropology, linguistics, semiotics, history, sociology, and others, featuring in structuralism, constructivism, and their modern successors.

In addition, philosophy of mind together with a conceptually inclined cognitive science (as opposed to experimental psychology) developed as offsprings, as it were, of the linguistic turn. The corresponding *cognitive turn* (Fuller *et al.*, 1989) redirected emphasis from language to cognition, and can be traced to the early cognitivism of Chomsky, Minsky and Simon. Today's implications of the cognitive turn are manifest in the study of consciousness, but also have visible repercussions in literature, theater, and film. This has recently led to the notion of an *iconic turn* (Maar and Burda, 2004), based on the idea that our interaction with the world essentially relies on images: classical images in the visual arts and in contemporary media, icons in communications with fellow humans and with computer systems.

This series of examples demonstrates how far remote present philosophical and cultural trends are from traditional metaphysics and ontology. It also shows the conjoining massive restriction of the scope of discourse from the quest for the fundaments of reality to language and cognition and eventually to visualization and its ramifications. In the resulting environment, a Cartesian substance dualism or the research programs of 19th century science must appear extremely naive. On the other hand, a narrow focus always makes it likely that important things outside of it are unduly disregarded. A comprehensive and sensible account of reality is palpably unachievable by elaborate studies of visual communication alone.

It is, therefore, easy to see that the ideas about reality that dominate contemporary science, humanities, and culture need to be considerably recast for an adequately shaped worldview. Such a recast may profit from reclaiming earlier ideas, but it also requires their reformation, rearrangement, and refinement. Ultimately, such an undertaking will be viable only if it proves successful. A specific difficulty in this respect is that new concepts and notions must be tried out, without established ways to test or apply them.

The life and work of Wolfgang Pauli, one of the leading theoretical physicists of the 20th century, offer illuminating and instructive material for corresponding studies. As Pauli wrote to Carl Gustav Jung at March 31, 1953 (Meyenn, 1999, p. 95), he was "baptized as 'anti-metaphysical' instead of Roman Catholic" due to the influence of his godfather Ernst Mach. So it is no surprise that Pauli belonged to the spiritual fathers of the operationally minded, or at least ontologically abstinent, Copenhagen interpretation of reality according to quantum mechanics.

So far, Pauli could simply appear as one of the early representatives of the trends sketched above. What makes his case particularly interesting, though, is his own "turn" back into metaphysics and ontology. This turn was initiated in the middle of his life, in the early 1930s, when he met the psychiatrist Jung at Zurich. Pauli adopted Jung's depth psychology rapidly and intensely. As a consequence, he started to develop and explore concepts going beyond his previous epistemic stance and tried to reconcile physics as a science of the material world with its non-material psychological counterpart. In a letter to Fierz of August 12, 1948, he wrote (Meyenn, 1993, p. 559):

> "When the layman says 'reality', he usually thinks that he is talking about something self-evident and well-known; whereas to me it appears to be the most important and exceedingly difficult task of our time to establish a new idea of reality."

Pauli addressed this issue only rarely in his regular publications, for example in his extensive essay on Kepler (Pauli, 1952), his article about central ideas of Jung's psychology (Pauli, 1954), and his historical account of Western science (Pauli, 1956). Nevertheless, as he expressed in a letter to Born of January 21, 1951, he saw his lasting impact beyond his achievements in physics in "the ideas that I communicate more or less directly to a small circle of scholars and friends" (Meyenn, 1996, p. 243). The main medium of communication for these ideas was his extraordinarily numerous correspondence in his letters. He used them mainly for two purposes: (i) to criticize work which he thought was wrong or, worse, not even wrong, and (ii) to discuss his speculative ideas beyond physics with colleagues. Pauli's complete correspondence has been so excellently edited by Karl von Meyenn[3] that it can now serve as an immensely rich source for studies of Pauli's extraphysical ideas.

This was one of the motives for a conference on *Wolfgang Pauli's Philosophical Ideas and Contemporary Science*, on which the present volume is based. The idea originated from a proposal by Ulrich Müller-Herold who, with his ingenious combination of persuasive and convincing talents, put together a board of organizers including himself, Karl von Meyenn, Reinhard Nesper, and the editors of this volume. He arranged that the conference could be held in May 2007 at the Centro Stefano Franscini at Monte Verità (Ascona, Switzerland), with both its splendid environment and its superb service. And, together with Reinhard Nesper and his staff, he made sure that all financial and administrational matters were lined up perfectly.[4]

Another reason for the conference, after an earlier predecessor in June 1993 at the same place (Atmanspacher *et al.*, 1995), was to relate Pauli's ideas to new developments in contemporary science and philosophy. The 1993 conference was held in cooperation with the Jung Institute Zurich and, accordingly, had a strong Jungian component. Since then, a number of prominent innovative developments related to Pauli's views occurred in fields other than Jung's psychology (see Atmanspacher and Primas, 2006). For this reason, Jungian perspectives were deliberately less considered, though not completely avoided, for the invitation of speakers for the 2007 conference.

Its main topics can be assigned to four areas: basic ideas in the philosophy of science and of mind, their relations to different notions of time, research about how creative insight operates, and new developments in biological evolutionary theory, especially epigenetics. Beyond those areas, there are two

---

[3] It contains more than 7000 pages in eight volumes, published over a quarter of a century between 1979 and 2005. Front-runners in the list of exchange partners are Heisenberg with 460 letters and Fierz with 350 letters. Jung and his circle are represented with 300 letters, and Bohr follows with 150 letters. Pauli's largely unpublished correspondence with Paul Rosbaud is estimated with 300 letters, but only few of them are presently accessible.

[4] The website at `http://www.solid.ethz.ch/pauli-conference/` provides some retrospective impressions of the conference and contains interesting photographs of Pauli, some of which are widely unknown.

contributions to this volume which may serve the reader as introductory material. First there is the article by *Karl von Meyenn* on the role that Pauli's correspondence plays for the study of his philosophical ideas. It addresses in detail Pauli's education in the positivist spirit of Mach and the Vienna circle, and then his departure from it.

In the second paper, *Domenico Giulini* gives an in-depth account of the role of symmetry principles in Pauli's work in physics. Fundamental symmetries were central in his thinking, and he warned against violating symmetry groups without good reasons.[5] This made his critical attitude in physics sometimes productive (e.g., prediction of the neutrino), but sometimes also obstructive (e.g., parity violation). It is interesting to see how symmetry principles also feature in Pauli's ideas beyond physics, for which Giulini indicates a pertinent example deserving further study.

The predominantly philosophical papers circulate around the idea of dual-aspect thinking and complementarity as a special variant thereof. *William Seager* presents an introduction to dual-aspect approaches as a combination of ontological monism with epistemological dualism. He traces this scheme back to Spinoza, where a self-contained *causa sui* creates many manifestations. Seager suggests that Pauli's ideas of mind and matter are much closer to Spinoza than this is visible in his writings. Spinoza's *causa sui* can be related to both Plato's archetypal *ideas* and to Jung's *unus mundus*, a basic form of reality of which the mental and the material are regarded as aspects.

Dual-aspect approaches to the mind-matter problem have been advertised again by physicists, for instance, by Bohm (1990) or by d'Espagnat (1999). None of them, however, has been worked out to an extent at which it leads beyond Pauli's or where it might even become operationally useful. A particularly promising feature of Pauli's dual aspects is their proposed complementarity.

Colloquially speaking, two descriptions of a situation are complementary if they are both necessary for a complete description of that situation and at the same time incompatible with each other. A precise characterization of complementarity as a logic with restricted sentential connectivity (which figures prominently in contemporary investigations under the name *partial Boolean algebras*) is due to Strauss (1936). It generalizes both classical and quantum logic and provides a formal basis to apply the concept of complementarity beyond quantum physics. In the present collection *Peter beim Graben and Harald Atmanspacher* show how this leads to deeper insight into the structure of epistemic descriptions of *classical* dynamical systems.

In the area of consciousness studies, established in the early 1990s, Chalmers (1996) proposed dual-aspect thinking as a way to address the "hard problem" of relating first-person and third-person accounts of consciousness to each other. Modifying Chalmers' approach, *Max Velmans* finds that complementarity offers a suitable framework for many of the properties that he

---

[5] See the letter of Pauli to Peierls of February 19, 1957 (Meyenn, 2005a, p. 244).

conceives as important. In his own “reflexive monism” he combines the reflexivity of phenomenal and neuronal aspects of a mind-brain with its ontically monistic, unified totality.

A specific example of complementarity applied to the mental domain is described in the article by *Harald Atmanspacher, Thomas Filk, and Hartmann Römer*. This example refers to a purely cognitive account (without invoking possible brain mechanisms as neural correlates) of the bistable perception of ambiguous stimuli. Based on the complementarity of the dynamics of spontaneous reversals between the two perceived states and the dynamics of observing those states, they present a formal model (the “Necker-Zeno model”) that is confirmed by a number of non-trivial experimental results. Their paper ends with the challenging proposal of a temporal variant of entanglement, a nonlocality in time, for unstable mental states.

Joachim Klose, in his contribution, reminds us of a non-mainstream philosophical approach which, nevertheless, has received increasing attention in recent years: Whitehead’s process philosophy. On Whitehead’s account, the basic elements of reality are “actual entities”, conceived similar to Leibniz’s monads, but in permanent interaction. Other than pointlike events in physical spacetime, actual entities are extended in space and time. They have both a mental and a physical pole, appearing as their coexisting aspects. This picture is central in the approach of Stapp (2007), which Klose discusses as a current attempt to use Whitehead for an interpretation of quantum theory including the mental.

Complementarity of mind and matter and the problem of time are the two basic topics that *Hans Primas* links to each other in his article. He proposes that mind and matter may be related via a temporal domain serving as an interface between atemporal material and mental domains. In the temporal domain, he distinguishes tenseless and tensed time, referring to the parameter time of physics and to our experiential distinction of past, present, and future, respectively. Primas understands these two concepts of time as contextual descriptive tools, emerging from an epistemic symmetry breaking of an underlying non-Boolean reality, the *unus mundus*. The mental and the physical arise as complementary and holistically correlated decompositions of this transcendental reality.

All these approaches, as different as they are in detail, reflect Pauli’s (1952) vision that “it would be most satisfactory if physis and psyche could be conceived as complementary aspects of the same reality”. Pauli speculated that the nature of this reality might have to do with the collective unconscious in the sense of Jung, without space and time and other categories with which the sciences of today operate. We know next to nothing about such a reality. Which symmetry of the *unus mundus* may be broken such that time emerges? Under which transformations would the description of an *unus mundus* be invariant, and how could such an invariance be detected? Why should a decomposition into tensed and tenseless, physical and mental domains be preferrable to others? Or, if it is not, what are other relevant decompositions?

These and more fundamental questions come to mind immediately, and they remain unanswered so far.

Ideas of the preceding contributions are taken up and merged in the article by *Georg Franck and Harald Atmanspacher*. If mind-matter relations can be rephrased in terms of relations between tensed and tenseless time, then the tension between the intensity of *mental presence* and the duration of the *temporal present*, of nowness, may be a key to further insight concerning the mind-matter problem. The authors outline some ideas of how cognitive time scales predicted by the Necker-Zeno model might indicate degrees of mental presence. Ultimately, this leads to the question where the most primordial forms of mental presence, or primary consciousness, begin: Some form of panpsychism is the price to be paid for the conceptual elegance of dual-aspect thinking, but maybe this price is just appropriate for the explanatory surplus to be gained.

Another feature of mind and matter as complementary aspects of a holistic *unus mundus* was proposed by Jung (1952) after long discussions with Pauli: synchronicity. *François Martin and Giuliana Galli Carminati* discuss synchronicity as an acausal (interaction-free) correlation between mental and physical states. They explain the seemingly paradoxical character of such correlations as a classical illusion comparable with delayed-choice experiments, where it *seems as if* results can be manipulated by changes of the past. Different from physical entanglement, it is a decisive feature of synchronistic relations that the correlated states share some subjectively experienced meaning. Martin and Galli Carminati outline a model of how meaningful emotional states can give rise to synchronistic effects between individuals.

If synchronistic correlations reflect the lost holism, or a broken symmetry, of the *unus mundus*, it becomes a pressing question how this fundamental reality can be conceived. Both Jung and Pauli speculated that basic elements of the collective unconscious, fundamental archetypes, might be interesting candidates in this respect. *Arthur Miller* elucidates this idea with an example from Pauli's biography. He recalls how Pauli comments his step from the three known degrees of freedom of the electron to a fourth, the electron spin, which led him to the formulation of the exclusion principle. From the viewpoint of his psychological development, Pauli interpreted this as a transition from a "trinitarian" to a "quaternarian" attitude, thus expressing the role of numbers as qualitative archetypal concepts (unity, duality, trinity, quaternity, ...) rather than tools for quantification.

Pauli – and with him other first-rate mathematicians like Hardy, Gödel, Penrose or Connes – looked for archetypal elements in the sense of Platonic ideas as the basis of mathematical truth. Since this Platonic conception of archetypes cannot be tested scientifically, *Rafael Núñez* suggests in his article to understand the foundations of mathematics as a product of the embodied human mind. He reinterprets a number of aspects of Jungian archetypes in terms of "image schemas", conceived as providing the link between cognition and language in contemporary cognitive neuroscience.

These different usages of the notion of archetype in different contexts might indicate a way to connect the situation shaped by linguistic and cognitive movements back to more ontological deliberations about the nature of reality. Any attempt at "recasting reality" must seriously take into account the present body of scientific knowledge and constructively use its results to refine earlier approaches. Pauli's vision of a "new idea of reality" strongly needs the substantial achievements of contemporary science (and the ability to distinguish them from the extraneous) for its realization.

The article by *Michael Öllinger and Günther Knoblich* is devoted to the psychology of insight, another contemporary topic of rapidly growing attention. The authors begin with an overview of early work by Gestalt psychologists such as Köhler, Wertheimer, and Duncker from the 1920s to the 1940s. Based on their results, different cognitive approaches have been developed subsequently, and the tedious experimental paradigms of current work show that the achieved understanding of insight progresses in very small steps. While current research on creative insight shows why solutions to difficult problems often occur suddenly and involuntarily, it cannot explain the intriguing creative experiences of a Gauss or a Poincaré, as described by Hadamard (1954). What made these men of genius so extraordinary (cf. Simonton, 1988, for corresponding ideas) is a question beyond those asked in ordinary insight research.

A particularly astonishing example is the Indian mathematician Ramanujan (1887–1920). With almost no training in mathematics and no access to mathematical libraries he had, at the age of 25, discovered and rediscovered more than 3000 mathematical theorems. After 30 years of studies of his notebooks all these theorems are now proven by methods unknown to himself, but the roads that led Ramanujan to his results have remained enigmatic for the most part. Ramanujan did not try to solve problems – he insisted that his insights were revealed to him by a family deity (see Kanigel, 1999).[6]

A further area of vivid interest to Pauli was biological evolution, addressed by *Linda van Speybroeck.* Pauli found that at least three critical issues were not sufficiently clarified by the standard neo-Darwinian picture of random mutations plus selection factors: Are the probabilities for the evolution of species estimated properly ? Are there environmental effects on genomes ? Is efficient causation enough to explain evolutionary mechanisms ? The first two questions are intensely studied in recent research on adaptive *non-random* mutations and on epigenetics, i.e. inheritable changes of phenotype *without* genotype changes. Only twenty years ago, such ideas were considered utmost heretical *vis-a-vis* the central dogmas of full-blown neo-Darwinism (see Jablonka and

---

[6] Skills with similarly mysterious origin are known in the context of the so-called *savant syndrome*, which gains increasing attention in current research. It refers to a rare condition in which persons with developmental disorders have one or more areas of expertise, ability or brilliance that are in contrast with the individual's overall limitations. For more details see Treffert (2006).

Lamb, 2005, for a review). Concerning the third question, however, there is still no evidence that teleology or final causes need to be involved, something that Pauli had in mind when he speculated about evolution as a series of meaningful events akin to synchronicity.

In correspondence with Delbrück, Weisskopf, Pittendrigh, Bohr and Elsasser, Pauli wrote about biological evolution in astonishing detail. And although Delbrück accused him of participating in a "plot of unemployed theoretical physicists against biology",[7] the recent development of genetics showed that Pauli's concerns were highly relevant. Additional unexplored territory, not so evident for him in his time, appears in connections between evolution and learning. One crucial point here is the riddle of the so-called major transitions in evolution (Maynard Smith and Szathmáry, 1995), resembling the phenomenon of punctuated equilibrium. Another one would be the role of epigenetic processes in neurons and associated progress in our understanding of neural plasticity based on Hebbian learning (Hebb, 1949).

In one way or another, the contents of this volume focus on new developments in philosophy and science in the light of Pauli's conjectures and speculations of more than half a century ago. In some cases, distinct progress is already visible, in others it can only be anticipated. Future generations of scholars will be able to see more clearly in which directions and with which understanding the concept of reality will develop. And they will be able to assess more distinctly the role which Pauli's expectations will play in this process:[8]

> "My personal opinion is that in a future science reality will be neither 'mental' nor 'physical' but somehow both of them and somehow neither of them. … Today both (micro-) physics and psychology (of the unconscious) deal with an *invisible* reality (or 'posit' such a reality, as philosophers say). As a consequence one has to be 'prepared' (old-Bohr-style) to find properties different from those of the macro-world."

References

Atmanspacher, H., Primas, H., and Wertenschlag-Birkhäuser, E. (eds.), (1995): *Der Pauli-Jung-Dialog und seine Bedeutung für die moderne Wissenschaft*. Springer, Berlin.

Atmanspacher, H., and Primas, H. (2006): Pauli's ideas on mind and matter in the context of contemporary science. *Journal of Consciousness Studies* **13**(3), 5–50.

Bohm, D. (1990): A new theory of the relationship of mind and matter. *Philosophical Psychology* **3**, 271–286.

Chalmers, D.J. (1996): *The Conscious Mind: In Search of a Fundamental Theory.* Oxford University Press, Oxford.

---

[7] Quoted from a letter of Pauli to Elsasser of September 30, 1958 (Meyenn, 2005b, p. 1271).

[8] Quoted from a letter of Pauli to Pais of August 17, 1950 (Meyenn, 1996, p. 152f).

d'Espagnat, B. (1999): Concepts of reality. In: Atmanspacher, H., Amann, A. and Müller-Herold, U. (eds.), *On Quanta, Mind, and Matter*. Kluwer, Dordrecht, pp. 249–270.

Fuller, S., de Mey, M., Shinn, T., and Woolgar, S. (eds.), (1989): *The Cognitive Turn: Sociological and Psychological Perspectives on Science*. Springer, Berlin.

Hadamard, J. (1954): *The Psychology of Invention in the Mathematical Field*. Dover, New York.

Hebb, D.O. (1949): *The Organization of Behavior*. Wiley, New York.

Jablonka, E. and Lamb, M.J. (2005): *Evolution in Four Dimensions*. MIT Press, Cambridge.

Jung, C.G. (1952): Synchronizität als Prinzip akausaler Zusammenhänge. In Jung, C.G. and Pauli, W. (eds.), *Naturerklärung und Psyche*, pp. 1–107. Rascher, Zürich.

Kanigel, R. (1991): *The Man Who Knew Infinity: A Life of the Genius Ramanujan*. Charles Scribners's Sons, New York.

Maar, C. and Burda, H. (eds.), (2004): *Iconic Turn. Die neue Macht der Bilder*. DuMont, Köln.

Maynard Smith, J. and Szathmáry, E. (1995): *The Major Transitions in Evolution*. Oxford University Press, Oxford.

Meyenn, K. von (1993): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band III: 1940–1949*. Springer, Berlin.

Meyenn, K. von (1996): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil I: 1950–1952*. Springer, Berlin.

Meyenn, K. von (1999): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil II: 1953–1954*. Springer, Berlin.

Meyenn, K. von (2005a): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil IV-A: 1957*. Berlin, Springer.

Meyenn, K. von (2005b): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil IV-B: 1958*. Berlin, Springer.

Pauli, W. (1952): Der Einfluss archetypischer Vorstellungen auf die Bildung naturwissenschaftlicher Theorien bei Kepler. In: Jung, C.G., and Pauli, W. (eds.), *Naturerklärung und Psyche*, pp. 109–194. Rascher, Zürich 1952. English translation in: C.P. Enz and K. von Meyenn (eds.), *Wolfgang Pauli. Writings on Physics and Philosophy*, Springer, Berlin 1994, pp. 219–279.

Pauli, W. (1954): Naturwissenschaftliche und erkenntnistheoretische Aspekte der Ideen vom Unbewussten. Dialectica **8**, 283–301. English translation in: C.P. Enz and K. von Meyenn (eds.), *Wolfgang Pauli. Writings on Physics and Philosophy*, Springer, Berlin 1994, pp. 149–164.

Pauli, W. (1956): Die Wissenschaft und das abendländische Denken. In: Göhring, M. (ed.), *Europa – Erbe und Aufgabe*. Steiner, Wiesbaden, pp. 71–79. English translation in: C.P. Enz and K. von Meyenn (eds.), *Wolfgang Pauli. Writings on Physics and Philosophy*, Springer, Berlin 1994, pp. 137–148.

Pauli, W. (1957): Phänomen und physikalische Realität. *Dialectica* **11**, 36–48. English translation in: C.P. Enz and K. von Meyenn (eds.), *Wolfgang Pauli. Writings on Physics and Philosophy*, Springer, Berlin 1994, pp. 127–135.

Petersen, A. (1963): The philosophy of Niels Bohr. *Bulletin of the Atomic Scientist* **19**(7), 8–14.

Rorty, R. (ed.), (1967): *The Linguistic Turn: Essays in Philosophical Method*. Chicago University Press, Chicago.

Simonton, D. (1988): *Scientific Genius. A Psychology of Science.* Cambridge University Press, Cambridge.

Stapp, H. (2007): *Mindful Universe*. Springer, Berlin.

Strauss, M. (1936): Zur Begründung der statistischen Transformationstheorie der Quantenphysik. *Sitzungsberichte der Preussischen Akadademie der Wissenschaften, physikalisch-mathematische Klasse* **27**, 382–398. English translation with an instructive postscript added in 1971 in: M. Strauss, *Modern Physics and Its Philosophy. Selected Papers in the Logic, History, and Philosophy of Science*, Reidel, Dordrecht 1972, pp. 186–203.

Treffert, D.A. (2006): *Extraordinary People. Understanding Savant Syndrome.* iUniverse, Lincoln.

Wolfgang Pauli's Philosophical Ideas Viewed from the Perspective of His Correspondence

Karl von Meyenn

Institute for Theoretical Physics, University of Ulm, Germany

Summary. Pauli grew up under the influence of Ernst Mach, but – like Einstein – he turned away from the radical positivism of most of his contemporaries quite early. Even though he was a rigorous and systematic thinker, he always devoted much attention to paradoxes and to the mystical background of science. Pauli tried to reconcile this attitude with both modern physics and Jung's archetypal psychology. While his publications present the results of more or less longsome searches for insight, his methodical flow of work and the gradual emergence of understanding become visible only in his rich correspondence.

1 The Traditional Relation between Physics and Philosophy

Relations between physics and philosophy have a long history, but a fundamental change in these relations occurred with the discovery of quanta.[1] Until then, it was considered the task of physics to identify rationally definable and empirically testable facts within the philosophically conceivable. It "is an attempt", so Markus Fierz, alluding to the famous prolog to Faust,[2] "to reconstruct the primordial images of appearances wavering in space." This illustrates how many of our classical notions were anticipated by philosophers, until they could – after proper transformation and adaptation to scientific demands – be completely incorporated into the domain of physics.

A particularly impressive example of such a conceptual development induced by philosophy is the often discussed, long history of the concept of the atom, which could ultimately be absorbed by physics only in recent times. During the 19th and still in the early 20th century renowned scholars such as Gustav Theodor Fechner, Ludwig Boltzmann, Wilhelm Ostwald, Ernst Mach,

---

[1] Cf. Arnold Sommerfeld's (1948) talk on "Physics and Philosophy" at an international summer school in Munich at July 3, 1948.

[2] Letter to Pauli, Meyenn (1996), p. XXXV, and Meyenn (1999), p. 636.

and Max Planck disputed questions of a “Physical and Philosophical Doctrine of the Atom”.[3] Only the experiments by Jean Perrin and their theoretical interpretation in the framework of statistical mechanics by Marian von Smoluchowski and Albert Einstein put a definitive end to such questions. The young Wolfgang Pauli could follow these problems at close range when he went to the “Döblinger Gymnasium” at Vienna. There he received his first scientific education under the supervision of his godfather Ernst Mach, who also advised Pauli to learn from the appropriate mathematics textbooks.

Physicists like Pauli were interested in the philosophy and historical origins of our modern scientific concepts early on. However, with the current publication style in the sciences the impact of such philosophical deliberations about the emergence of new ideas is usually not focused at explicitly. For this reason it is mostly very difficult to fathom the role that philosophical questions play in the development of scientific ideas. Because Pauli belonged to those physicists who were raised in the tradition of writing letters, his case puts us into the beneficial situation that we can close this gap of knowledge to a considerable extent.

Pauli always looked for the company of philosophically and historically educated colleagues when he tried to learn about the state of the art of their research. If the person he wanted to talk to was difficult to approach, he often decided to make a detour via their collaborators who were supposed to launch his request in the right moment. In this way, Pauli was able to keep continuous contact to Bohr and Jung even in heavy-duty periods. Because these conversations were often accompanied by correspondence, we possess – particularly for the later years – revealing evidence for his philosophical ideas.

Pauli studied numerous philosophical and other publications by well-known scholars and authors which are partly conserved and can be accessed in the library assembled at CERN in Geneva. During reading he sometimes annotated passages that he found remarkable or objectionable by bulky marks or brief notes. This provides important indications for his thinking, which are also useful for a better understanding of his letters. A comprehensive reprint collection, including publications of more general content, supplements the rich source material that is now available for contemporary historians of science.[4]

2 Physical Concepts Without Philosophical Precedents: Quantum Physics and Pauli’s Exclusion Principle

The role of philosophy as a source of ideas rapidly terminated with modern physics. With the more and more boosting art of experimentation physicists

---

[3] This was the title of a monograph published in 1855 (second edition 1864) by the Leipzig physicist Gustav Theodor Fechner (1801–1887), founder of the field of psychophysics.

[4] See the overview of Pauli’s estate in Section 10 of this contribution.

at the turn of the past century entered realms beyond our daily experience and discovered that both the microcosm and the macrocosm host phenomena that can no longer be mapped onto patterns accessible by our sensory organs. In particular, the realm of the atoms, and the quantum theory describing them, required completely novel conceptions, impossible to find by resorting to existing philosophical approaches.

In his Nobel lecture at Stockholm, Pauli (1946) described the shock "which every physicist, accustomed to the classical way of thinking, experienced when he came to know of Bohr's 'basic postulate of quantum theory' for the first time." Then he continued to report how he himself managed to overcome this crisis with the help of the already existing work of his two teachers (see Enz and Meyenn, 1994, p. 166):

> "At that time there were two approaches to the difficult problems connected with the quantum of action. One was an effort to bring abstract order to the new ideas by looking for a key to translate classical mechanics and electrodynamics into quantum language which would form a logical generalization of these. This was the direction which was taken by Bohr's Correspondence Principle. Sommerfeld, however, preferred, in view of the difficulties which blocked the use of the concepts of kinematical models, a direct interpretation as independent of models as possible, of the laws of spectra in terms of integral numbers, following, as Kepler once did in his investigation of the planetary system, an inner feeling for harmony. Both methods, which did not appear to me irreconcilable, influenced me."

Pauli's own contribution to the foundations of the new quantum theory was his somewhat Pythagorean-like *exclusion principle*,[5] assigning spin as a classically not existing property to a particular class of elementary particles, which Pauli presented as "antisocial particles" in his Nobel address. Although the exclusion principle, also called "Pauli Verbot",[6] was still formulated in the framework of the semi-classical Bohr-Sommerfeld quantum theory, its fundamentally non-dynamic character was not clear until Heisenberg discovered the anti-symmetry of the wave function in summer 1926.[7] The requirement of anti-symmetry introduces a novel kind of correlation between electrons that was alien to physics so far. It implies that two particles must neither come to close to each other nor travel with speeds that are too similar to each other.

In a letter of January 24, 1927, Paul Ehrenfest posed the witty question of "whether in recent times hardly a topic proves viable if it did not first receive the blessings of the curse of the Pauli-Verbot ? Eventually, the ennobling accolade is hardly anything else than a stylish slap in the face" (Hermann *et al.*, 1979, p. 372). Pauli's discovery had such a fundamental significance for the behavior of atomistically constituted matter that the Russian physicist Hans

---

[5] This characterization is due to Einstein in a letter to his friend Michele Besso of November 30, 1949.

[6] A literal translation would be "Pauli proscription".

[7] A particularly instructive example for the impact of anti-symmetrization on the behavior of two electrons is discussed in Chapter 20 of Margenau (1950).

Hellmann claimed in his “Introduction to Quantum Chemistry” (Hellmann, 1937): “Everything non-classical ultimately follows from the Pauli principle and the existence of a kinetic zero-point energy of the electrons.” A number of phenomena unexplained so far could suddenly be understood: the structure of atoms and molecules, the nature of directed valence forces, ferro- and paramagnetism, the spectra of atoms and molecules, and the stability of matter in general.

Acknowledging the outstanding importance of the discovery of the exclusion principle, the Dutch physicists honored Pauli in 1931 with the Lorentz medal, endowed just a year before. In his humorous address, Ehrenfest (1931, p. 621) could not resist to allude to the pitiless criticism with which Pauli did not even spare his best friends (as the attending Bohr):

> “Sometimes you do even accomplish that your closest and most trusted friend impatiently jumps out of his otherwise carefully balanced vocabulary and syntax. ... Yes, Mr. Pauli, finally you will not succeed to restrain all your contemporaries from appreciating you very highly, even adore you, and thus wish you all the best for your work and for your personal bliss.”

In the new edition of his Handbuch article “Philosophy of Mathematics and Natural Science”, Hermann Weyl (1949) had claimed that Leibniz’s identity principle was a classical precursor of the Pauli principle. Weyl explicitly emphasized that “the Leibniz–Pauli exclusion principle [holds] for electrons but not for photons” (Weyl, 1949, p. 247). Pauli immediately raised vehement objections. Supported by his colleague and philosophical advisor Markus Fierz at Basel he convinced Weyl that his claim was untenable: “A philosophical principle, like the ‘principium identitatis indiscernibilium’, should after all not be understood such that it holds for some object but not for others” (Meyenn, 1993, p. 701).

Further investigations of quantum phenomena revealed additional cases of non-classical behavior, among them the tunnel effect, the indistinguishability of elementary particles (described as “Selbigkeit” by Schrödinger), the fundamentally statistical character of Schrödinger’s wave function, and finally the violation of parity, which troubled Pauli until to the last years of his life.

Given all these new developments, Pauli now posed to philosophers the inverse task to augment their concepts and adapt them to the improved body of knowledge in atomic physics. In a letter to the philosopher Hermann Levin Goldschmidt (1990, p. 41) of February 19, 1949, he wrote:[8]

> “It seems to me as a philosophical layman that the task of philosophy consists in generalizing the emerging insights of current physics – that is, all its essential elements – in such a way that it can be applied to fields more general than physics. Such an achievement would, in turn, enrich the individual disciplines and prepare future developments.”

---

[8] Hermann Levin Goldschmidt (born 1914) had visited a lecture by Pauli at February 8, 1948, and sent him his book *Philosophie der Dialogik* (Goldschmidt, 1948) immediately the next day. Pauli’s letter is a reaction to this book.

Later, in his "Mainz sermon" (see Section 9), Pauli indicated how he thought about this in more detail.

3 Pauli, Mach, and Positivism

Positivism was the philosophical guidance system of scientific research, based on classical physics, in the early 20th century (cf. Frank, 1917). It is an attempt to organize scientific progress in terms of clear-cut recipes. Theories were supposed, as Pauli recalled during a philosophy congress in Zürich in summer 1954, "to be derivable by compelling logical conclusions from minute books" (Pauli, 1957, p. 38; see also Kraft, 1950, pp. 108ff; Holton, 1973, p. 145).

Ernst Mach, who had given the seventh edition of his famous "Mechanics" as a gift to Pauli in 1913, also advised Pauli's father in the education of the precocious boy. As the widow Franca Pauli reported, he was quite hot-tempered, and in one of his outbreaks of displeasure he had even smashed his mother's valuable Chinese vase. Only after he discovered mathematics and its wonderful presentation in Leonhard Euler's "Introductio In Analysis Infinitorum" (printed 1748 in Lausanne) and other ambitious mathematical *opera*, the world turned less inane for him. Letters of his father and reports of contemporaries such as the Vienna physicist Hans Thirring testify that Pauli had the reputation of a mathematical genius already in his school days at the Gymnasium in Vienna.[9]

In the fifth chapter of his "Mechanics" Mach discussed the "Relations of Mechanics to Other Domains of Knowledge". Here Pauli highlighted a paragraph essentially outlining the positivist program:

> "The mechanistic world view seems to us as a historically understandable, excusable, maybe even temporarily useful, yet on the whole artificial hypothesis. If we want to remain faithful to the method that led most important scientists such as Galilei, Newton, Carnot, Faraday, Mayer to their

---

[9] In a letter of August 8, 1914, to Wilhelm Jerusalem, philosopher at the University of Vienna since 1892, Ernst Mach writes from Haar/Munich: "Profesor Pauli spent a few days here with his son. He believes that he is a profound mathematical genius."

Hans Thirring recalled in a broadcast address at December 19, 1958, that Pauli already as an adolescent showed such extraordinary talent that he was described as a child prodigy "who – as Mozart – met all the expectations. ... During the first world war, 1915 or 1916, a younger colleague of mine, teaching at the Gymnasium in the XIVth quarter, told me one day: 'Imagine, in the fifth class we have a schoolboy with such a phenomenal talent for mathematics and physics that he promises to become a new Gauss or Boltzmann.' "

When the young Pauli studied with Sommerfeld in Munich, Sommerfeld sometimes asked Pauli for advice to resolve mathematical difficulties. At some occasion Weyl had submitted a new mathematical treatment of a problem to Sommerfeld, whereupon the latter complained (in a letter of January 6, 1920) that "I myself, but even Pauli, had major difficulties to follow your discussion."

> great discoveries, we restrict our physics to the expression of the factual and stay away from hypotheses about anything behind the factual, where nothing is tangible or testable. Then we need to simply determine the real relation between motions of masses, changes of temperature, variations of the values of potential functions, chemical rearrangements, without assuming anything else underneath these elements than physical features or characteristics given to us indirectly or directly by observation."

Later Pauli dismissed this methodology, also recommended by neo-positivists, as too one-sided and emphasized that a "creative irrational element" is involved when something novel is being found. He saw a promising access to our understanding of the process of scientific discovery in Jung's psychology of the unconscious, which he began to look into besides his purely scientific research.

4 The Article on Relativity: Felix Klein Introduces Pauli to the Art of Scientific Writing

Already at school Pauli had, supported by the Vienna lecturer Hans Bauer (1891–1953), gotten access to tensor calculus. It was difficult to learn for a schoolboy but inevitable to understand the then new general theory of relativity. So it happened that Pauli in his first semester at the University of Munich surprised his teacher Sommerfeld with two finished contributions to relativity, which even aroused Einstein's attention. This led Sommerfeld to entrust an article on relativity to Pauli, which was to complete the volume on mechanics within the "Encyclopedia of Mathematical Sciences".[10]

While other students were occupied with their lectures and exercises, Pauli used the first two years of his study to write this article. He received particular support by the great mathematician Felix Klein in Göttingen. As a founder of the Erlangen program, he was one of the pioneers of relativistic physics. He introduced Pauli "not only into the subject but also into the art of disposition and scientific style."[11] In addition, Klein provided lecture manuscripts and other notes for Pauli's work and advised him far beyond usual measures. On April 20, 1920, Klein informed Einstein:[12] "Luckily, at the moment work on the mathematical encyclopedia is making better progress again. In particular, we are approaching relativity theory from astronomical and physical angles

---

[10] Originally, Einstein was commissioned to write this article. "As Einstein declined this offer", Sommerfeld said when he recommended Pauli as a corresponding member of the Bavarian Academy of Sciences in 1948, "I proposed to Pauli to write it together with me. But when he showed me his first drafts, I abandoned the idea of a joint project. His article became a masterpiece that is unmatched until today."

[11] Quoted after a contribution by Wilhelm Wirtinger, Vienna, to the Festschrift for Klein published in 1919.

[12] Buchwald *et al.*, 2004, p. 535.

(Kottler under the supervision of Oppenheim, Pauli under the supervision of Sommerfeld)."

When the 250-pages and 400-footnotes article, finished in December 1920, appeared in print at November 15, 1921, Pauli was already a scientific celebrity. The proficient editor Arnold Berliner was even afraid that Pauli might become megalomaniac because of Einstein's overwhelming appraisals. But the latter could appease him with the remark that this premonition came too late.

By his collaboration with Klein, Pauli had become acquainted with the most esteemed scientists of his time. And he became familiar with the mathematical tools that were exquisitely suitable for dealing with the upcoming problems of theoretical physics. As hardly anyone else he was equally familiar with relativity theory and quantum theory, the two most demanding fields of theoretical physics. So he was ideally prepared for the challenges that physical research had in store for the coming decades.

5 Moritz Schlick and the "Vienna Confession"

In an early correspondence with Moritz Schlick, the leading philosopher of the "Vienna Circle" who in 1922 was appointed the chair formerly held by Mach, Pauli evinced his interest for epistemology and natural philosophy.[13] Here he expresses his philosophical inclination for the very first time. On August 15, 1922, Schlick had sent him the fourth edition of his book on "Space and Time in Contemporary Physics" with thanks for the "hours spent in Rostock" together. At this meeting, to which Pauli came from Hamburg, their conversation apparently led into diverging opinions concerning a publication by Joseph Petzoldt, an adherent of Mach. Pauli asserted (Meyenn, 1985, p. 692) that he had "looked carefully into Schlick's objections against positivism once again" and could "no longer acknowledge them as sound." Underlining his personal conviction, Pauli emphasized once more that he thought of "positivism as a completely coherent world view, free of contradictions", even though obviously "not the only one possible".

A few years after the new quantum mechanics was established, Pauli received a programmatic publication from the "Vienna Confession", just founded by Moritz Schlick, Rudolf Carnap, Hans Hahn, Otto Neurath and

---

[13] Pauli knew well that Einstein also held Schlick in high esteem, both as a philosopher and as a physicist. Einstein had conveyed to Schlick on December 14, 1915, that he thought of his publications as "among the best that has been written about relativity so far." He added "you [also] saw correctly that this line of thought had a great influence on my efforts, and more specifically, E. Mach, and even more so Hume, whose 'Treatise of Human Nature' I had studied avidly and with admiration shortly before discovering the theory of relativity" (Schulmann *et al.*, 1998, p. 220).

Ludwig Wittgenstein. He still found it quite interesting "but I do not feel entirely affiliated with it" any longer.[14]

During the 1920s Schlick had repeatedly commented on the position of the principle of causality within physics; however, he almost exclusively concerned himself, somehat one-sidedly, with the consequences brought about by the theory of relativity. Later he turned to quantum theory (Schlick, 1931, p. 145):

> "But now that the viability of quantum theoretical concepts is confirmed by the extraordinary success of its applications, and we had quite a few years to get used to the new ideas, now the attempt should no longer be premature to achieve philosophical clarity concerning the meaning and the impact of the thoughts that current physics contributes to the problem of causality."

This agreeable statement notwithstanding, Schlick failed to address the idea of complementarity, so fundamental for quantum mechanics. Moreover, Pauli criticized Schlick's imprecise formulation (Meyenn, 1985, p. 56): "The point is that I can interpret everything you say in such a way that I agree. However, much can also be interpreted such that I had to protest. Briefly, I think you did not express yourself precise and clear enough in all the questions you raise." This might be the origin of the popular Pauli quote "This is not even wrong!"

When Pauli visited the USA in summer 1931 and in winter 1935/36, he met a number of emigrants who now established a closer relationship with him. During a trip to Chicago he got to know the physicist Carl Henry Eckart, a friend of Carnap. Eckart had made important contributions to the development of wave mechanics. Moreover, he had helped to translate Heisenberg's lectures on "The Physical Principles of Quantum Theory" at the University of Chicago, which stimulated his epistemological interests. Recently, some exchange of letters with Pauli surfaced in Eckart's estate. These letters give us new insights into Pauli's philosophical views and general interests during the 1930s.

In a letter of January 17, 1936, Pauli asked Eckart for his "further spiritual and human relation to the Vienna confession", which he was still attached to (as another letter of February 11, 1936, shows). At the same time he couched

---

[14] Meyenn, 1985, p. 15; see also Geier (1992). Pauli's library included Rudolf Carnap's programmatic volume "Der logische Aufbau der Welt" (Carnap, 1928) which Pauli had carefully read and annotated. Concerning Carnap's demand (preface, p. V) "to dispel all of metaphysics from philosophy, because its hypotheses cannot be rationally justified" and "every scientific thesis must be rationally substantiated", Pauli noted: "The fact that science is done at all cannot be rationally justified!"

Pauli's aversion against an absolutistic attitude with respect to philosophical systems was primarily directed against Kant's dogmatic *a priori* conditions. He reinforced this in his letter to Goldschmidt (1990, p. 39) of February 19, 1949: "Rational ideas are never necessary or certain and always object to rational criticism. No rational idea resides in an unassailable olympus of necessities of thought."

his critical stance concerning symbolic logic, which was then much discussed in positivist circles. In particular he saw a restricted role for mathematics to play in future physics:

> "The symbolic logic has, according to my opinion, not a direct applicability to physics, because the theoretical physics represents physics by mathematics ('bildet die Physik auf die Mathematik ab'). So the symbols involved are mathematical symbols and their connection with each other is a mathematical question. (I agree on this point completely with what you quote as Dirac's opinion.) – But the main difference of mathematics and physics is the connection of the mathematical symbols (or at least some of them) with empirical results – that means in the last end with some sensations which are made artificially simple. And in this latter connection all logical paradoxes or antinomies of the human knowledge come into play. One of them concerns the notions of subject and object and consists in the fact that on the one side it is necessary to distinguish between a recognizing subject and a recognized content in order to be able to formulate any knowledge; that on the other side every content of thoughts is also a part of the subjects. Both sides of the situation of human knowledge are equally important and the best we can do is to put them on the beginning as necessary conceptual antinomies (not 'paradoxes')."

6 Departing from Positivism: Complementarity, C.G. Jung, and the Problem of Opposites

The epistemological shifts that accompanied the discovery of quantum mechanics and its interpretation were partly responsible for Pauli's altered view on positivism. But before we go into details, let us give a general overview of Pauli's philosophical development as he saw it himself:[15]

> "What impressed me philosophically at all, I can ... only indicate very briefly: opposite Mach (empiricism) – Plato (ideas at 'heavenly location'), Kant (the preconditions for the natural sciences of his time are dogmatically fixed and erroneously considered as the quintessential preconditions of human reason, the *a priori* is ascribed to rationally formulated ideas) – modern psychology of the 'unconscious' (Freud, C.G. Jung) (the *a priori* lies in preconscious states – *esse in anima* – 'archetype' as pathway for imagination = pre-existing images as in Plato, Proclus, Kepler). Then: enlightenment (Voltaire, Mach) – on the other hand Vedanta teachings, Schopenhauer ('will' as his God). (P.S. Bernard Shaw's remark that unmasking a heavenly 'Hauptmann von Köpenick' does not prove that a real 'Hauptmann' exists, as I noted.) The entire East impressed me strongly. China much more than India, both the ideas of the I-Ging (Yin-Yang-polarity) and Laotse. Schopenhauer's attempt to reconcile Kant and Buddhism seemed very interesting to me but remained unsuccessful as a consequence of Kant's backwardness and Buddha's passivity vis-a-vis the world. In general the 17th

---

[15] Quoted again from the letter to Goldschmidt (1990, pp. 29–31) of February 19, 1949.

> century (besides much more ancient times) means a lot to me and the 19th century little. German intellectuality always appeared to me to tend towards dogmatism and kinds of one-sidedness that are foreign to the instincts. How different are the wise men of China! And everything collective-crowdlike is much afar from my taste in general. Furthermore, it seems to me that feeling is as deep as thinking and that *amo ergo sum* would be as justified as the *cogito ergo sum* by Avicenna–Descartes.[16] (P.S. Pathological exaggeration of the thinking function by Hegel.) In this atmosphere, looking for a balance within pairs of opposites, I grew up from the earliest days of my boyhood."

As one of the founders of the new quantum theory Pauli belonged to the most fervent advocates of the so-called Copenhagen interpretation. In September 1927 he retreated to "Villa Monte Pensada" close to Como together with Bohr for joint discussions of the notion of complementarity.[17] This notion, originally introduced by Bohr, was thought to enable a synthesis of the seemingly contradictory dualism of wave and particle.[18] It turned out that for this purpose extensions of the usual notions of causality and reality were needed. In the quoted letter to Goldschmidt (1990, p. 37) Pauli specified the epistemological significance of complementarity:

> "However, the modern physicist does not refer to a 'complementary' situation as contradictory but he characterizes his description (since 1927) as contradiction-free (English: 'self-consistent'). The range of applications of opposing concrete images (such as 'wave' and 'particle') in the new theory is now delineated in such a way that contradictions cannot occur. What appears are no 'contradictions' but is rather a limitation of the applicability of our ways of perception, not only by the possibilities of observation but also by the possibilities of definition (caused by the laws of nature)."

Later on, attempts have been made to apply complementarity also to problems outside physics, e.g. a complementarity of clarity and truth (Pauli's letter to Goldschmidt, 1990, p. 33):

> "If a proposition is too clear, then something goes wrong with its correctness, and if a proposition is true, then its clarity is limited. For every truth contains something partly unknown, only foreboded, and thus also a hidden opposite of its conscious meaning."

Pauli tried to illustrate the complementary distinction between symbolic and quantitative descriptions with the schema shown in Tables 1 and 2.

Pauli's epistemological conceptions reveal the influence of Jung's psychology of the unconscious, with which he concerned himself, also scientifically, since his marriage with Franca in April 1934. For instance, observations of

---

[16] Pauli used these comparisons also in his "Philosophical Comedy" of 1952, see Meyenn 1996, pp. 464, 493.

[17] This information is due to an interview with Oskar Klein of February 28, 1963.

[18] A clear exposition of this problem area can be found in Pauli (1950).

dreams were considered as options to track processes of the unconscious. During his psychoanalysis Pauli had learned how to decode the language of his dreams; now he wanted to continue this activity out of scientific curiosity. By the technique of *amplification*, the contents of private dreams could be related to and interpreted by events of both own experiences and – in agreement with the idea of a collective unconscious – myths of ancient times or foreign cultures.

It was important for Pauli's efforts to decipher the language of dreams and other manifestations of the unconscious that it can be comprehended only indirectly and *symbolically.*[19] Pauli regarded the quantum mechanical ψ-function, which relates possible observational data to each other (as a probability amplitude), and adopts the role of Kant's things-in-themselves as such a symbol uniting opposites. In the letter to Goldschmidt (1990, p. 39) of February 19, 1949 he explains:

> "The symbol is always an abstract token, be it quantitative or qualitative, be it mathematical-theoretical or emotionally laden. Only part of the symbol can be expressed by conscious concepts, another part acts on the 'unconscious' or 'preconscious' state of an individual. The same holds for mathematical symbols, for only he is gifted for mathematics for whom mathematical tokens (in the sense mentioned above) have symbolic power. The symbol always is a tertium uniting opposites, what logic alone cannot 'provide'."

Pauli considered it as a remarkable coincidence whenever novel concepts appeared simultaneously in completely different areas, e.g. the introduction of the notion of a physical field and the discovery of the unconscious in psychology.[20] As the electromagnetic field

> "was theoretically related to a reality, no matter whether or not it can be visualized by suitable means, the unconscious was related to a reality as an edge layer of subliminal 'contents' which, however, can possibly influence consciously perceived processes considerably."

According to Freud, this "subliminal something, somehow controlling consciousness from behind the scene", was based on "contents repressed from consciousness". Jung, on the other hand, attributed it also to "collective contents which had never been conscious before".

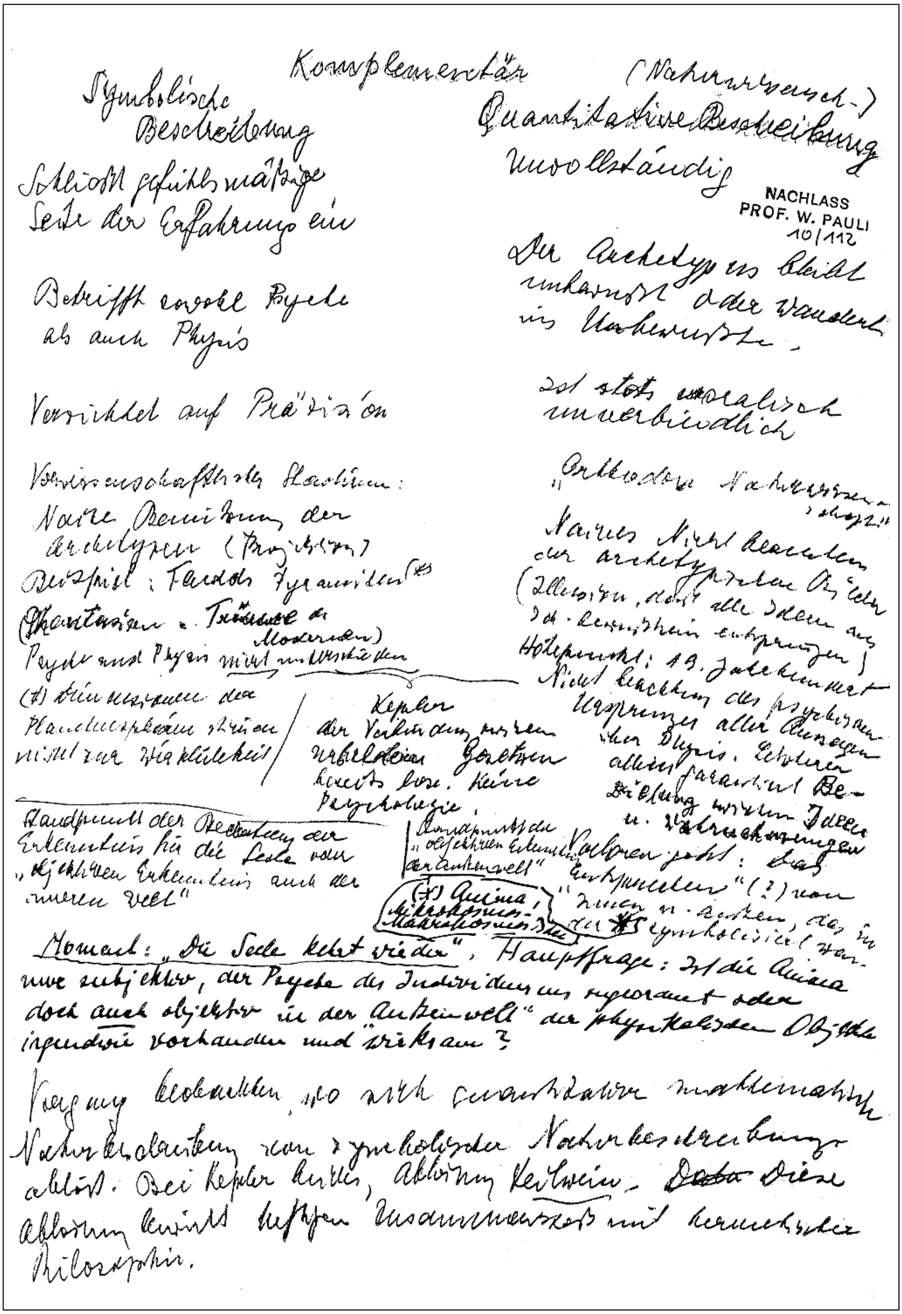
For a while Pauli was so fascinated by the interpretation of dreams that some of his friends started to demur. When the mathematician Erich Hecke at Hamburg heard about Pauli's visit to Princeton in fall 1935, he wrote in a letter to Weyl of October 31:

> "Probably you took Pauli with you when you traveled back. His wife, whom I find very cute, hopefully accompanied him. Yes, he depends very much on

---

[19] Compare Pauli's notes on "Modern Examples of Background Physics" which he comprised for Jung in summer 1948 (Meier, 2001, pp. 179–196).

[20] See Pauli (1954), p. 283. He made similar remarks in a letter to C.A. Meier of February 26, 1950 (see Meyenn, 1996, pp. 35ff).



Komplementär

Symbolische Beschreibung

(Naturwissensch.) Quantitative Beschreibung

Schließt gefühlsmäßige Seite der Erfahrung ein

unvollständig

NACHLASS PROF. W. PAULI 10/112

Betrifft sowohl Psyche als auch Physis

Der Archetypus bleibt unbewußt oder wandert ins Unbewußte.

Verzichtet auf Präzision

Ist stets moralisch unverbindlich

Vorwissenschaftliches Stadium: Naive Benützung der Archetypen (Projektion) Beispiel: Fludds Pyramiden(*) Phantasien u. Träume d. Modernen) Psyche und Physis nicht unterschieden

„Orthodoxe Naturwissenschaft"

Naives Nichtbeachten der archetypischen Bilder (Illusion, daß alle Ideen aus dem Ich-Bewußtsein entspringen) Höhepunkt: 19. Jahrhundert

(*) Dimensionen der Planetensphären stimmen nicht zur Wirklichkeit

Kepler der Verbindung mit den urbildlichen Gedanken bereits bewußt. Keine Psychologie.

Nichtbeachten des psychischen Ursprungs aller Aussagen über Physis. Erfahrung allein garantiert Beziehung zwischen Ideen u. Wahrnehmungen

Standpunkt der Beachtung der Erkenntnis für die Seele oder „objektiven Erkenntnis auch der inneren Welt"

Standpunkt der „objektiven Erkenntnis der Außenwelt"

(*) Anima, Mikrokosmos-Makrokosmos-Idee

Naturen jetzt: das „Entsprechen" (?) von Innen u. Außen, das in der ✱ symbolischen Vor-

Motto: „Die Seele kehrt wieder". Hauptfrage: Ist die Anima nur subjektiv, der Psyche des Individuums zugeordnet oder doch auch objektiv in der „Außenwelt" der physikalischen Objekte irgendwie vorhanden und wirksam?

Vorgang beobachten, wo sich quantitative mathematische Naturbeschreibung von symbolischer Naturbeschreibung ablöst. Bei Kepler erstes, Ablösung teilweise. ~~Dabei~~ Diese Ablösung bewirkt tiefen Zusammenstoß mit hermetischer Philosophie.

Table 1. Handwritten outline on "complementary modes of description" by Pauli (document 10/112 in the Pauli estate at CERN Geneva)

<table>
<tr><th colspan="2">complementary</th></tr>
<tr><th>symbolic description</th><th>quantitative description
(natural science)</th></tr>
<tr><td>includes emotional side of experience</td><td>incomplete</td></tr>
<tr><td>concerns both mental and physical aspects</td><td>the archetype remains unconscious or moves into the unconscious</td></tr>
<tr><td>abstains from precision</td><td>always morally noncommittal</td></tr>
<tr><td>pre-scientific phase:
naive use of archetypes (projection)</td><td>“orthodox natural science”
naive ignorance of archetypal images (illusion that all images arise from ego-consciousness)</td></tr>
<tr><td>example: Fludd's pyramids*)
phantasies and dreams of modern man
*) dimensions of planetary spheres do not agree with reality</td><td>apex: 19th century</td></tr>
<tr><td>mental and physical <i>not distinguished</i></td><td>disregarding the mental origin of all propositions about the physical; only the latter guarantees relations between ideas and perceptions</td></tr>
<tr><td colspan="2">Kepler
the connection between primordial images and laws is already loose;
no psychology;
attitude of “objective knowledge of the external world”</td></tr>
<tr><td>attitude of the significance of knowledge for the soul or “objective knowledge of the inner world as well”</td><td>lost: the “correspondence” (?) of inside and outside, symbolized by the anima, idea of microcosm–macrocosm</td></tr>
<tr><td colspan="2"><i>instant: “the soul returns”</i>, main question: is the amina only subjective, associated with the psyche of individuals, or <i>also</i> objectively existing and efficacious in the “external world” of physical objects?
Study the process by which a quantitative mathematical description of nature separates from a symbolic description of nature. Both present in Kepler, <i>partial</i> separation, causing severe clash with hermetic philosophy.</td></tr>
</table>

Table 2. Translated reconstruction of Table 1

> supporting help, this silly billy, with all his extraordinary intelligence. What I came to know about him most recently concerning his actual condition is really unedifying. For years he is now under treatment by Jung because his nerves caused him great trouble. Lastly he was so obsessed by his treatment that he talked about nothing else than his dreams, and daily affairs played a role for him only insofar as they were reflected by his dreams. This is a huge piece of work for his wife."

In spite of this psychological pressure Pauli was able to do important research in physics in those years. And he did not only observe and analyze his dreams – he also drew far-reaching consequences as to the role of the unconscious in the evolution of science. In the mentioned letter to Eckart, Pauli referred to his novel views on scientific creativity, as they had changed due to Jung's psychology:

> "A similar antinomy arises from the concepts of 'consciousness' and the 'Unconscious' – the latter as an idea being on the other hand also a content of the consciousness (ein 'Bewusstseinsinhalt'). I would like to make the statement that every concept (Begriff) describing our knowledge can by analysis in the last end be reduced to such not further analyzable antinomies (and just if it would be otherwise, then it would be something wrong with the underlying concepts.) – It seems to me that the connection between symbols and experience cannot be enlightened by symbols again because those would remain always on the one side. There must be some place where the individual 'Hinweise' to concrete objects come into play.
> What we only can do is to show how human knowledge and particular sciences, as physics for instance, *do really proceed*. And then we shall not find confirmed the desires (Wünsche) of individual philosophers and philosophical systems. We shall find neither the pure inductive nor the pure deductive type of physics possible and we shall find sometimes that first the empirical results were present and after that one has found the symbolic mathematical description of them, and sometimes also the opposite was the case.
> I personally have, besides, not much interest to fix the state of any science in some accidental point of time axiomatically, but merely to look in what direction a further development of this science is possible. (And so, I think, the most satisfactory situation is, if the axiomatics would always come too late.)"

7 Princeton, Panofsky and the Kepler Article

During the war, when Pauli lived in the USA for an extended period of time, he entertained some epistemologically oriented correspondence with the philosopher Hans Reichenbach, who had emigrated to California in 1938. When Reichenbach had finalized the manuscript for his book "Philosophical Foundations of Quantum Mechanics" (Reichenbach, 1944), he asked Pauli for advice concerning "causal anomalies".[21] Reichenbach's proposal of a three-valued logic in this context did not find Pauli's support. On the other hand he did not want to argue against Reichenbach's concepts "in the sense of an anti-metaphysical vice squad."[22] "As a physicist I prefer to leave the laws of logic and the axioms of mathematics untouched as a sound basis." Although Pauli was familiar with the problematic crisis of axiomatic foundations, he recommended that physicists should act according to the dictum *divide et impera*.[23]

The art historian Erwin Panofsky, who had been a young reader at the University of Hamburg at the same time as Pauli, was of major influence for Pauli's further career. He was a member of the Humanities Department at the Institute for Advanced Study (IAS) at Princeton since 1935.[24] The stimulation that Pauli received through his contacts with Panofsky and other members of this department, such as the philologist Harold F. Cherniss and the historian Ernst Kantorowicz, sparked his interest in Renaissance philosophy, which eventually led to the Kepler article published in a joint book with Jung in 1952. Inspired by his dream analysis, unveiling his transformation from a *trinitarian* to a *quaternarian* attitude, Pauli intended to illustrate the impact of Jungian *archetypes* and the role of the *collective unconscious* with the example of Kepler.

The idea of such a study apparently originated at the IAS Princeton with its excellent library[25] and a circle of scholars who were open to interdisciplinary topics.[26] In addition to Panofsky, Cherniss and Kantorowicz, it was mainly Max Knoll, the co-inventor of the electron microscope, who stayed at

---

[21] Reichenbach (1948) authored a contribution entitled "The Principle of Anomaly in Quantum Mechanics" for the issue of the journal *Dialectica* that was edited by Pauli.

[22] Quoted from a letter of Pauli to Reichenbach of January 6, 1943, which will be published in the supplement volume to Pauli's correspondence edition.

[23] Pauli explained his position in the letter to Eckart of February 29, 1936: "My opinion is that logic and mathematics are different in their content ('Inhalt') more than in their form. In mathematics one wants to derive from given axioms new concepts and new consequences. And I think that the particular choice of axioms which is done in mathematics is not accidental. Further I think that just these particular axioms of mathematics are suited to give a scientific description of nature as it does physics."

[24] Compare Meyenn, 2005, p. 237. Panofsky's extensive correspondence is being edited by Dieter Wuttke.

[25] At February 26, 1950, Pauli reports from Princeton, full of enthusiasm, "that another colleague at the Humanities Department owns an original version of Fludd's 'Philosophia Moysaica' (it is supposed to be the only copy available in the USA)" (Meyenn, 1996, p. 35).

[26] Pauli's occupation with Kepler is first mentioned in his letter to Fierz of December 29, 1947 (Meyenn, 1993, pp. 488, 496).

Princeton together with his wife Ursula[27] and was very much interested in *synchronistic phenomena*.[28]

In his study, Pauli related the rise of the modern scientific world view to a repression of religious feelings exclusively into domains of the church. This, he claimed, was accompanied by a transition from a quaternarian to a trinitarian attitude which took place in both collective and personal realms of the psyche (Meyenn, 1993, p. 706):

> "For this reason, it is important even today to reformulate the principle of synchronicity as a further principle for the explanation of nature, on equal footing with and independent of causality, i.e. complementing it, in a suitable way. Only such an explanation of nature could be called quaternarian, while present-day physics is still trinitarian."

8 Collaboration with Philosophers at Zurich: "What Went Where?"

After his return to Zurich in spring 1946 Pauli established contacts with the philosophers at his university. In particular, he made friends with the Austrian-Hungarian philosopher Franz Kröner (1889–1958) who had studied physics and mathematics, and later philosophy, at Vienna and joined the Polytechnicum at Zurich as a scientific assistant to Ferdinand Gonseth in 1951 (Meyenn, 1999, p. 111). Pauli became a frequent visitor of the history-of-science seminars run by Gonseth and Paul Bernays.

Moreover, he served on the advisory board of the journal "Dialectica" published by Swiss philosophers. He also helped to organize several philosophical conferences which he animated sanguinely with sketchy formulations, for instance referring to meetings of "Knights at the Round Table". In 1948 a special issue of "Dialectica" was published on the idea of complementarity under the patronage of Pauli, with contributions by Einstein, Bohr, Heisenberg, Reichenbach, and de Broglie. In a lecture at the "Philosophical Society" Zurich in 1949 Pauli indicated the general possibilities which the idea of complementarity, grown out of atomic physics, holds for a reintegration of a science that has fragmented into many subdisciplines. He deplored that "in contrast to the theory of relativity, this turn in modern physics has been realized only by a small number of philosophical specialists" (Pauli, 1950, p. 72).

Under the impression of the progress of quantum mechanics, Pauli had more and more distanced himself from his earlier positivist stance. However,

---

[27] Compare Meyenn, 1996, pp. 55f. Knoll (1952) gave a lecture on "Wandlungen der Wissenschaft in unserer Zeit" at the Eranos Meeting 1951.

[28] See Meyenn, 1993, pp. 706f. Jung denoted phenomena as synchronistic if they are connected by a common meaning but have no physical explanation (Jung, 1952, p. 83). The notion is derived from Leibniz's parable of synchronized clocks for the illustration of mind-matter relations.

some physicists regarded the development of quantum physics as a strict consequence of the positivist program. Pauli, on the other hand, emphasized that the epistemological situation of modern physics "was not anticipated by any philosophical system". He liked to scoff at the tendency of philosophers toward systematization and was noncommittal with respect to any of the philosophical schools ending with *...ism.*[29] While philosophers like to subordinate their entire thinking under a system, physicists rather tend to be more eclectic. Depending on circumstances they borrow ideas from different philosophical systems and do not care much about philosophical vicissitude. It is this rather positive sense in which one has to interpret Pauli's statement that Fermi was a semi-empirical opportunist because he did not systematically develop his theory of β-decay from first principles.

Anyway, Pauli himself did "not intend to become a founder of religion or philosophy with advancing age", such as Bohr for instance,[30] "who decidedly has a tendency to perform as the originator of a 'religion of complementarity'. My stance is rather to find some balance between extreme directions", he declared during the philosophy congress at Zurich in 1954.

Much in the spirit of psychological practice Pauli observed subtle changes of historical background. He considered it particularly meaningful when certain concepts disappeared and were replaced by others. Using the example of the vanished concept of freedom in a Cartesian world view, he commented (Meyenn 1996, p. 472):

> "Even if one does not share the naive belief in progress of the 19th century, it is very instructive to investigate the history of ideas – and the history of physics and the sciences as well – from the viewpoint: *What went where?* For we learned that every act of conscious realization is paid for by the fact that something which was conscious beforehand – even though sometimes vaguely – falls back into the unconscious and may reappear 'in an altered shape' as a revenant."

In a letter to von Weizsäcker he prompted him to "rewrite the history of ideas and of science from the perspective of the persisting question: What went were?" (Meyenn, 1999, p. 142).

9 Science and Occidental Thinking

In fall 1954 Pauli had read "a book on West-Eastern mysticism and another one about telepathy" in preparation for an upcoming congress at Mainz.[31] In an elaborate letter to Jung's secretary Aniela Jaffé he outlined his preliminary

---

[29] Compare the correspondence with Kröner with some examples of Pauli's derogatory remarks on the idiosyncrasies of philosophers.

[30] Quoted from a letter to Heisenberg of May 13, 1954 (Meyenn, 1999, p. 620).

[31] Compare the commentary by Meyenn (1999), pp. 629f, and a folder labeled "Mainzer Vortrag 1955" and "Unity of Knowledge von Bohr" in Pauli's estate.

ideas with respect to his congress contribution on "Science and Occidental Thinking". He conceived the interrelation of mystic experience and rational understanding in the evolution of occidental thinking "as a being awake that is a dream, and a dreaming that is like being awake." After leaving aside this work for a while, he turned back to it early in 1955.

In mid January 1955 he finished a first draft of the text, which now was only to be transcribed and "polished". "The most difficult thing was that the talk should only take 45 minutes (with an extensive discussion afterwards). However, what I think about the problem of how redemptive knowledge [Heilserkenntnis] and scientific knowledge are related to each other comes out quite well now." Pauli sent a copy to Kröner, "partly for checking historical details". In mid February he asked Kröner for information about other participants and traveled to Mainz at March 16 "to sing his song to an unknown crowd" (see letter to Panofsky in Meyenn, 2001, p. 154).

In a compact historical overview of the "problem of the relation between redemptive knowledge and scientific knowledge" Pauli argued that "periods of dispassionate research on critical lines are often succeeded by others in which the aim is to try to include science in a more comprehensive spiritualism involving mystical elements." Finally he makes the far-reaching statement that (see Enz and Meyenn, 1994, p. 147)

> " . . . at the present time a point has again been reached at which the rationalist outlook has passed its zenith, and is found to be too narrow. Externally all contrasts appear to be extraordinarily accentuated. On one hand the rational way of thought leads to the assumption of a reality which cannot be directly apprehended by the senses, but which is comprehensible by means of mathematical or other symbols, as for instance the atom or the unconscious. But on the other hand the visible effects of this abstract reality are as concrete as atomic explosions, and are by no means necessarily good, indeed sometimes the extreme opposite. A flight from the merely rational, in which the will to power is never quite absent as a background, to its opposite, for example to a Christian or Buddhist mysticism is obvious and is emotionally understandable. Yet I believe that there is no other course for anyone for whom narrow rationalism has lost its force of conviction, and for whom also the magic of a mystical attitude, experiencing the external world in its crowding multiplicity as illusory, is not effective enough, than to expose himself in one way or another to these accentuated contrasts and their conflicts."

Again Pauli presents a union of opposites as a goal, a kind of theory of everything, in which rational understanding and a mystical experience of unity in the sense of Bohr's idea of complementarity are to be reconciled.

After his return to Zurich he communicated his general impression about the "current spiritual situation in the occident" to Panofsky (Meyenn, 2001, p. 196):

> "In Mainz I realized that *the evil* (inquisition, combats of sects, communism – in my opinion a Christian sect with "matter" as its superior metaphys-

> ical principle with the status of a goddess) is not sufficiently accepted as *occidental.* There are gentlemen who carry an "iron curtain" (namely of repression) *in themselves.* The symmetry between inside and outside seems perturbed."

Pauli was satisfied with the success of his lecture. He appreciated in particular the acquaintance with the historian of science Willy Hartner from Frankfurt, former collaborator of the sinologist Richard Wilhelm whom he admired much. All lectures were published in a volume entitled "Europe – Heritage and Challenge", edited by the director of the Department for Universal History of the Mainz Institute for European History. After Pauli had received and corrected the page proofs of his text (Pauli, 1956), the volume appeared early in 1956. On the occasion of a visit to Hamburg at the end of November 1955 he repeated his lecture for a different audience in the Jungius Society.

When Fierz proposed to him to "compose a broadly conceived historical-critical study reaching up until present times" under the title "Thoughts and Background Ideas of a Modern Physicist", Pauli thought seriously about it. Such a study would have created great interest among physicists and non-physicists. However, he did not get around to working on it. As a consequence of the new developments accompanying the discovery of the neutrino, Pauli turned back to physical problems during the last years of his life and postponed his more private interests.

10 Overview of Pauli's Scientific Estate

Pauli published more than 200 articles and essays in both German and English language, most of which are reprinted in the two-volume edition of his "Collected Scientific Papers" (Kronig and Weisskopf, 1964). His 1921 article on relativity, his two "Springer Handbuch" articles on the old and the new quantum theory, and his two contributions on radiation theory and atomic theory in "Müller-Pouillet's Lehrbuch der Physik" belong to the classics of physics literature, which served as textbooks for generations of physicists.

Possibly even more impact on the development of theoretical physics had his letters, with which he intervened into ongoing research in an influential way. These letters played an important role in the formation of opinions, were often shown around and willingly collected and conserved because of their contents and incisive formulations. After Pauli's death, his widow recollected many letters with the help of Bohr and some of Pauli's assistants, in order to edit and supply them for historical research. Presently, the published subset of his correspondence comprises about 3500 letters from and to Pauli, which are available for research on 7500 printed pages in eight volumes.[32] Comparing

---

[32] A little less than half of the letters are letters to Pauli. An additional supplement volume with further 400 letters, manuscripts and various tables and registers, which are to serve a facilitated use of the complete edition, is in preparation.

this amount with the 2500 pages of his published papers provides a rough idea of the influence of the letters on their receivers.

Those letters which Pauli possessed when he was still alive and those which were collected by his widow after his death are now deposited in the Pauli Archive at CERN in Geneva. The archive contains a collection of more than 10.000 reprints, a small library as well as notes, memoranda and manuscripts from Pauli's estate, which can be accessed via the world wide web at `library.cern.ch/archives/pauli/paulimain.html`.

A larger number of letters, particularly from the properties of Fierz and Jung and his coworkers are preserved in the history of science collections of ETH Zurich. Other comprehensive collections of Pauli letters are stored in the Niels-Bohr-Institute at Copenhagen and the Werner-Heisenberg-Institute (Max-Planck-Institute for Physics) at Munich. The remaining correspondence is scattered over various archives worldwide and could only be discovered with the help of directories and electronic databases that are available for historians today.

The major part of the letters is of physical content. Because many of them, in particular from the period before and during the war, have been lost, the current inventory provides a somewhat distorted picture of the actual extent of the correspondence with individual correspondents. Nevertheless, the high percentage of letters exchanged with Heisenberg (460 letters, 15%), Fierz (350 letters, 10%) and Bohr (150 letters, 5%) demonstrates the role of those physicists for Pauli's thinking.

Taking into account that Pauli sometimes contacted his correspondents through their close collaborators yields a considerable amount of 300 letters for the psychological correspondence with Jung. Another special case is Pauli's correspondence with ETH Zurich (Enz *et al.*, 1997) which sheds some light on several otherwise enigmatic aspects of Pauli's biography. Moreover, the correspondence with Paul Rosbaud, about 300 letters of which only a few have been made available so far, might be of mainly biographical interest.

Acknowledgment

I am grateful to Harald Atmanspacher for the translation of my German manuscript into English.

References

Buchwald, D.K., Schulmann, R., Illy, J., Kennefick, D.J. and Sauer, T. (eds.) (2004): *The Collected Papers of Albert Einstein. Volume 9. The Berlin Years: Correspondence, January 1919–April 1920.* Princeton University Press, Princeton. English translations in: Hentschel, A.M. (transl.) (2004), Princeton University Press, Princeton.

Carnap, R. (1928): *Der logische Aufbau der Welt.* Weltkreis-Verlag, Berlin.

Ehrenfest, P. (1931): Ansprache zur Verleihung der Lorentzmedaille an Professor Wolfgang Pauli am 31. Oktober 1931. In: Klein, M.J. (ed.), *Paul Ehrenfest. Collected Scientific Papers.* North-Holland, Amsterdam, 1959.

Enz, C.P., Glaus, B. and Oberkofler, G. (1997): *Wolfgang Pauli und sein Wirken an der ETH Zürich.* vdf Hochschulverlag an der ETH, Zürich.

Enz, C.P. and Meyenn, K. von (eds.) (1994): *Wolfgang Pauli. Writings on Physics and Philosophy.* Springer, Berlin.

Frank, P. (1917): Die Bedeutung der physikalischen Erkenntnistheorie Machs für das Geistesleben der Gegenwart. *Naturwissenschaften* **5**, 65–72.

Geier, M. (1992): *Der Wiener Kreis: mit Selbstzeugnissen und Bilddokumenten.* Rowohlt, Hamburg.

Goldschmidt, H.L. (1948): *Philosophie als Dialogik.* Aehren Verlag, Affoltern a.A.

Goldschmidt, H.L. (1990): *Nochmals Dialogik.* ETH Stiftung Dialogik, Zürich.

Hellmann, H. (1937): *Einführung in die Quantenchemie.* Deuticke, Leipzig.

Hermann, A., Meyenn, K. von and Weisskopf, V.F. (eds.) (1979): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel. Band I: 1919–1929.* Springer, New York.

Holton, G. (1973): *Thematic Origins of Scientific Thought: Kepler to Einstein.* Harvard University Press, Cambridge (Massachusetts).

Jung, C.G. (1952): Synchronizität als ein Prinzip akausaler Zusammenhänge. In: Jung, C. G., and Pauli, W. (eds.), *Naturerklärung und Psyche.* Rascher Verlag, Zürich, pp. 1–107.

Margenau, H. (1950): *The Nature of Physical Reality.* McGraw-Hill, New York.

Meier, C.A. (2001): *Atom and Archetype. The Pauli/Jung letters, 1932–1958.* Princeton University Press, Princeton.

Meyenn, K. von, (ed.) (1985): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band II: 1930–1939.* Springer, Berlin.

Meyenn, K. von, (ed.) (1993): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band III: 1940–1949.* Springer, Berlin.

Meyenn, K. von, (ed.) (1996): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil I: 1950–1952.* Springer, Berlin.

Meyenn, K. von (ed.) (1999): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil II: 1953–1954.* Springer, Berlin.

Meyenn, K. von, (ed.) (2001): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil III: 1955–1956.* Springer, Berlin.

Meyenn, K. von, (ed.) (2005): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil IV-A: 1957.* Springer, Berlin.

Knoll, M. (1952): Wandlungen der Wissenschaft in unserer Zeit. In: Fröbe-Kapteyn, O. (ed.), *Eranos-Jahrbuch 1951. Band XX. Mensch und Zeit.* Rhein-Verlag, Zürich, pp. 387–436.

Kraft, V. (1950): *Der Wiener Kreis: der Ursprung des Neopositivismus.* Springer, Wien.

Kronig, R. and Weisskopf, V.F. (eds.) (1964): *Collected Scientific Papers by Wolfgang Pauli. In Two Volumes.* Interscience, New York.

Pauli, W. (1950): Die philosophische Bedeutung der Idee der Komplementarität. *Experientia* **6**, 72–81.

Pauli, W. (1954): Naturwissenschaftliche und erkenntnistheoretische Aspekte der Ideen vom Unbewussten. *Dialectica* **8**, 283–301.

Pauli, W. (1956): Die Wissenschaft und das abendländische Denken. In: Göhring, M. (ed.), *Europa – Erbe und Aufgabe*. Steiner, Wiesbaden, pp. 71–79.

Pauli, W. (1957): Phänomen und physikalische Realität. *Dialectica* **11**, 36–48.

Reichenbach, H. (1944): *Philosophical Foundations of Quantum Mechanics*. University of California Press, Berkeley.

Reichenbach, H. (1948): The principle of anomaly in quantum mechanics. *Dialectica* **2**, 337–350.

Schlick, M. (1931): Die Kausalität in der gegenwärtigen Physik. *Naturwissenschaften* **19**, 145–162.

Schulmann, R., Kox, A.J., Jansen, M. and Illy, J. (eds.) (1998): *The Collected Papers of Albert Einstein. Volume 8A. The Berlin Years: Correspondence, 1914–1917*. Princeton University Press, Princeton. English translations in: Hentschel, A.M. (transl.) (1998), Princeton University Press, Princeton.

Sommerfeld, A. (1948): Philosophie und Physik seit 1900. In: *Arnold Sommerfeld. Gesammelte Schriften. Band IV*. Vieweg, Braunschweig, 1968, pp. 640–643.

Weyl, H. (1949): *Philosophy of Mathematics and Natural Science*. Princeton University Press, Princeton.

Concepts of Symmetry in the Work of Wolfgang Pauli

Domenico Giulini

Max-Planck-Institute for Gravitational Physics (Albert-Einstein-Institute),
Golm/Potsdam, Germany, `domenico.giulini@aei.mpg.de`

Summary. Symmetry was one of the most important methodological themes in 20th-century physics and is probably going to play no lesser role in the physics of the 21st century. As used today, there are a variety of interpretations of this term, which differ in meaning as well as their mathematical consequences. Symmetries of crystals, for example, generally express another kind of invariance than gauge symmetries, though in specific situations the distinctions may become quite subtle. I will review some of the various notions of symmetry and highlight some of their uses in specific examples taken from Pauli's scientific œvre.

1 General Introduction

In the Introduction to Pauli's Collected Scientific Papers, the editors, Ralph Kronig and Victor Weisskopf (1964, Vol. 1, p. viii), make the following statement:

> "It is always hard to look for a leading principle in the work of a great man, in particular if his work covers all fundamental problems of physics. Pauli's work has one common denominator: his striving for symmetry and invariance. ... The tendency towards invariant formulations of physical laws, initiated by Einstein, has become the style of theoretical physics in our days, upheld and developed by Pauli during all his life by example, stimulation, and criticism. For Pauli, the invariants in physics where the symbols of ultimate truth which must be attained by penetrating through the accidental details of things. The search for symmetry and general validity transcends the limits of physics in Pauli's work; it penetrated his thinking and striving throughout all phases of his life, in all fields of philosophy and psychology."

Those of Pauli's scientific contributions, which make essential use of symmetry concepts and applied group theory, certainly include the following, which form a substantial part of Pauli's scientific œvre:[1]

---

[1] Two of the listed themes, "meson-nucleon interaction and differential geometry" and "unifying non-linear spinor equation", were never published in scien-

> Relativity theory and Weyl's extension thereof (1918–1921), hydrogen atom in matrix mechanics (1925), exclusion principle (1925), anomalous Zeeman effect and electron spin (1925), non-relativistic wave-equation for spinning electrons (1927), covariant quantum electrodynamics (1928, with Jordan), neutrino hypothesis (1930), Kaluza-Klein theory and its projective formulation (1933), theory of γ-matrices (1935), Poincaré-invariant wave equations (1939, with Fierz), general particle statistics and Lorentz invariance (1940, with Belinfante), spin-statistics connection (1940), once more general relativity and Kaluza-Klein theory (1943, with Einstein), meson-nucleon interaction and differential geometry (1953), charge-parity-time symmetry (1955), β-decay and conservation of lepton charge ("Pauli group", 1957), unifying non-linear spinor equation (collaboration with Heisenberg, 1957–1958), group structure of elementary particles (1958, with Touschek).

Among the theoretical physicists of his generation, Pauli was certainly outstanding in his clear grasp of mathematical notions and methods. He had a particularly sober judgement of their powers as well as their limitations in applications to physics and other sciences. Let us once more cite Kronig and Weisskopf (1964, Vol. 1, p. viii):

> "Pauli's works are distinguished by their mathematical rigor and by a thorough and honest appraisal of the validity of assumptions and conclusions. He was a true disciple of Sommerfeld in his clear mathematical craftsmanship. By example and sharp criticism he constantly tried to maintain a similarly high standard in the work of other theoretical physicists. He was often called the living conscience of theoretical physicists."

It seems plausible that this critical impregnation dates back to his school days, when young Pauli read, for example, Ernst Mach's critical analysis of the historical development of the science of mechanics, a copy of which Pauli received as a present from his godfather Mach at around the age of fourteen. Mach's "Mechanik", as this book is commonly called, starts out with a discussion of Archimedes' law of the lever, thereby criticizing the following symmetry consideration (Mach, 1933, pp. 11–12): Imagine two equal masses, M, and a perfectly stiff and homogeneous rod of length L, both being immersed into a static homogeneous vertical gravitational field, where the rod is suspended at its midpoint, m, from a point p above (see Fig. 1). What happens if we attach the two equal masses to the ends of the rod and release them simultaneously without initial velocity ?

An immediate symmetry argument suggests that it stays horizontal; it might be given as follows: Everything just depends on the initial geometry

---

tific journals (on the second topic Heisenberg published results of a collaboration with Pauli without Pauli's consent) but can be followed from his letters and manuscripts as presented in Hermann *et al.* (1979), Meyenn (1985), Meyenn (1993), Meyenn (1996), Meyenn (1999), Meyenn (2001), Meyenn (2005a), Meyenn (2005b).

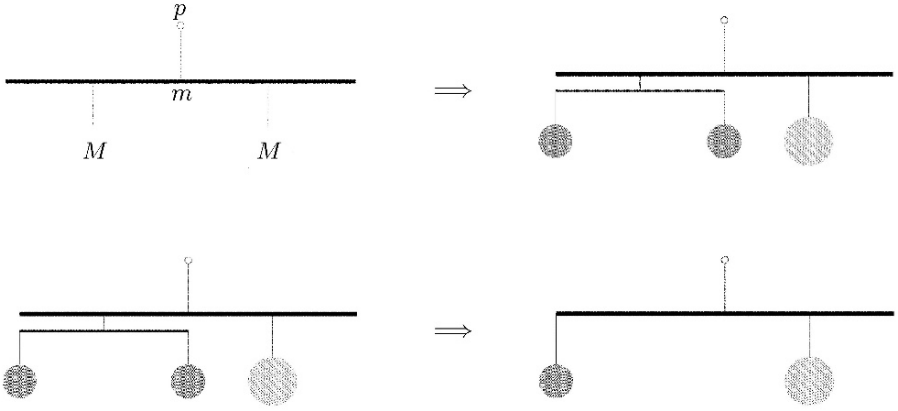


Fig. 1. The law of the lever "derived" from alleged symmetry considerations. The small balls represent half the mass of the big balls. The step from the upper right (second) to the lower left (third) picture does not follow.

and distribution of masses, which is preserved by a reflection at the plane perpendicular to the rod through p and m. Suppose that after release the rod dropped at one side of the suspension point m, then the mirror image of that process would have the same initial condition with the rod dropping to the other side. This is a contradiction if the laws governing the process are assumed to be reflection symmetric and deterministic (unique outcome for given initial conditions).

This argument seems rigorous and correct. Now, how does one get from here to the law of the lever? The argument criticized by Mach is as follows: Assume that the condition for equilibrium depends only on the amount of mass and its suspension point on the rod, but not on its shape. Then we may replace the mass to the left of m by two masses of half the amount each on a small rod in equilibrium, as shown in the second (upper right) picture. Then replace the suspension of the small rod by two strings attached to the left arm of the original rod, as shown in the third (lower left) picture, and observe that the right one is just under the suspension point m, so that it does not disturb the equilibrium if it were cut away as in the last (lower right) picture of Fig. 1.

The weak point in the argument is clearly the transition from the second to the third picture: There is no *global* symmetry connecting them, even though locally, i.e. regarding the small rod only, it connects two equilibrium positions. It is easy to see that, in fact, the assumption that a global equilibrium is maintained in this change is *equivalent* to Archimedes' law of the lever. This example shows (admittedly in a fairly trivial fashion) that alleged symmetry properties can work as a *petitio principii* for the law to be derived. This is essentially the criticism of Mach.

The reason why we consider this "derivation" of the law of the lever to be a *petitio principii* is that we have other, physically much more direct ways to properly derive it from *dynamical* first principles. From that point of view the alleged symmetry is to be regarded as an artifact of the particular law and certainly not *vice versa.* The observed symmetry requires an explanation in terms of the dynamical laws, which themselves are to be established in an independent fashion. This is how we look upon, say, the symmetry of crystals or the symmetric shape of planetary orbits.

On the other hand, all fundamental dynamical theories of 20th century physics are motivated by symmetry requirements. They are commonly looked at as particularly simple realizations of the symmetries in question, given certain *a priori* assumptions. It is clear that, compared to the previous example, there are different concepts of symmetry invoked here. However, there also seems to be a shift in attitude towards a more abstract understanding of "physical laws" in general.

What makes Pauli an interesting figure in this context is that this shift in attitude can be traced in his writings. Consider special relativity as an example, thereby neglecting gravity. One may ask: What is the general relation between the particular symmetry (encoded by the Poincaré group) of spacetime and that very same symmetry of the fundamental interactions (weak, strong, and electromagnetic, but not gravity) ? Is one to be considered as logically prior to the other ? For example, if we take Einstein's original operationalist attitude, we would say that the geometry of spacetime is defined through the behavior of "rods" and "clocks", which eventually should be thought of as physical systems obeying the fundamental dynamical laws. In fact, Einstein often complained about the fact that rods and clocks are introduced as if they were logically independent of the dynamical laws, e.g., in a discussion remark at the 86th meeting of the "Gesellschaft Deutscher Naturforscher und Ärzte" in Bad Nauheim in 1920: [2]

> "It is a logical shortcoming of the theory of relativity in its present form to be forced to introduce measuring rods and clocks separately instead of being able to construct them as solutions to differential equations."

From that viewpoint, symmetry properties of spacetime are nothing but an effective codification of the symmetries of the fundamental laws. Consequences like length contraction and time dilation in special relativity are then only *effectively* described as due to the geometry of spacetime, whereas a fundamental explanation clearly has to refer to the dynamical laws that govern clocks and rods. This was clearly the attitude taken by Lorentz and Poincaré, though in their case still somehow afflicted with the idea of a material ether that, in principle, defines a preferred rest frame, so that the apparent validity

---

[2] German original (Einstein, 1920): "Es ist eine logische Schwäche der Relativitätstheorie in ihrem heutigen Zustande, daß sie Maßstäbe und Uhren gesondert einführen muß, statt sie als Lösungen von Differentialgleichungen konstruieren zu können."

of the principle of relativity must be interpreted as due to a "dynamical conspiracy".[3] In his famous article on relativity for the "Encyclopedia of Mathematical Sciences", the young Pauli proposes to maintain this view, albeit without the idea of a material ether. He writes (Pauli, 1958a, p. 15):[4]

> "Should one, then, ... completely abandon any attempt to explain the Lorentz contraction atomistically? We think that the answer to this question should be 'no'. The contraction of a measuring rod is not an elementary but a very complicated process. It would not take place except for the covariance with respect to the Lorentz group of the basic equations of electron theory, as well as those laws, as yet unknown to us, which determine the cohesion of the electron itself. We can only postulate that this is so, knowing that then the theory will be capable of explaining atomistically the behavior of moving measuring rods and clocks."

Very recently, this traditional view has once more been defended under the name of "physical relativity" (Brown, 2005) against today's more popular view, according to which special relativity is about the symmetry properties of spacetime itself. Clearly, the latter view only makes sense if spacetime is endowed with its own ontological status, independently of the presence of rods and clocks.

This shift in emphasis towards a more abstract point of view is also reflected in Pauli's writings, for example in the preface to the English edition of his "Theory of Relativity" of 1956, where the abstract group-theoretic properties of dynamical laws are given an autonomous status in the explanation of phenomena:

> "The concept of the state of motion of the 'luminiferous ether', as the hypothetical medium was called earlier, had to be given up, not only because it turned out to be unobservable, but because it became superfluous as an element of a mathematical formalism, the group-theoretical properties of which would only be disturbed by it. By the widening of the transformation group in general relativity the idea of a distinguished inertial coordinate system could also be eliminated by Einstein, being inconsistent with the group-theoretical properties of the theory."

---

[3] Lorentz still expressed this viewpoint well after the formulation of special relativity (see, e.g., Lorentz, 1914, p. 23).

[4] German original (Pauli, 2000b, p. 30): "Ist aber das Bestreben, die Lorentz-Kontraktion atomistisch zu verstehen, vollkommen zu verwerfen? Wir glauben diese Frage verneinen zu müssen. Die Kontraktion des Maßstabes ist kein elementarer, sondern ein sehr verwickelter Prozeß. Sie würde nicht eintreten, wenn nicht schon die Grundgleichungen der Elektronentheorie sowie die uns noch unbekannten Gesetze, welche den Zusammenhalt des Elektrons selbst bestimmen, gegenüber der Lorentz-Gruppe kovariant wären. Wir müssen eben postulieren, daß dies der Fall ist, wissen aber auch, daß dann, wenn dies zutrifft, die Theorie imstande sein wird, das Verhalten von bewegten Maßstäben und Uhren atomistisch zu erklären."

Pushed to an extreme, this attitude results in the belief that the most fundamental laws of physics are nothing but realizations of basic symmetries. Usually this is further qualified by adding that these realizations are the most "simple" ones, at least with respect to some intuitive measure of simplicity. Such statements are well known from Einstein's later scientific period and also from Heisenberg's "unified theory" of elementary particles, for which he proposed a single non-linear differential equation whose structure was almost entirely motivated by its symmetry properties. Heisenberg made this point quite explicitly in his talk entitled "Planck's discovery and the foundational issues of atomism",[5] delivered during the celebrations of Max Planck's 100th birthday (at which occasion Pauli received the Max-Planck medal *in absentia*), where he also talked about his "unified theory":[6]

> "The mentioned equation contains, next to the three natural units $[c, \hbar, l]$, merely mathematical symmetry requirements. These requirements seem to determine everything else. In fact, one should just regard this equation as a particularly simple representation of the symmetry requirements, which form the actual core of the theory."

Pauli, who briefly collaborated with Heisenberg on this project, did not at all share Heisenberg's optimism that a consistent quantum field theory could be based on Heisenberg's non-linear field equation. His objections concerned several serious technical aspects, overlayed with an increasing overall dislike of Heisenberg's readiness to make premature claims, particularly when made publicly.

However, it is fair to say that the overall attitude regarding the heuristic role and power of symmetry principles in fundamental physics, expressed by Heisenberg in the above quote, was also to a large extent shared by Pauli, not only in his later scientific life. This is particularly true for symmetry-induced conservation laws, towards which Pauli had very strong feelings indeed. Examples from his later years will be discussed in later sections (e.g. Sect. 3.7). An example from his early scientific life is his strong resistance against giving up energy-momentum conservation for individual elementary processes, while keeping it on the statistical average. Such ideas were advocated in the "new radiation theory" of Bohr, Kramers, and Slater (1924), and again by Bohr in connection with β-decay, which Pauli called "spiritual somersaults" in a letter to Max Delbrück. A week after his famous postcard suggesting the existence

---

[5] German original: "Die Plancksche Entdeckung und die philosophischen Grundfragen der Atomlehre".

[6] German original (Meyenn, 2005b, p. 1168): "Die erwähnte Gleichung enthält neben den drei natürlichen Maßeinheiten nur noch mathematische Symmetrieforderungen. Durch diese Forderungen scheint alles weitere bestimmt zu sein. Man muß eigentlich die Gleichung nur als eine besonders einfache Darstellung der Symmetrieforderungen, aber diese Forderungen als den eigentlichen Kern der Theorie betrachten."

of the neutrino, Pauli wrote to Oskar Klein in a letter dated December 12, 1930:[7]

> "First it seems to me, that the conservation law for energy-momentum is largely analogous to that for electric charge, and I cannot see a theoretical reason why the latter should still be valid (as we know empirically from β-decay) if the former fails. Secondly, something strange should happen to the *weight* if energy conservation fails ... This contradicts my physical intuition to an extreme! For then one has to even assume that the gravitational field *produced* ... by the box (including the radioactive content) can change, whereas the electrostatic field must remain unchanged due to charge conservation (both fields seem to me analogous; as you will remember from your five-dimensional past)."

This is a truly remarkable statement. Not many physicists would nowadays dare suggesting such an intimate connection between the conservation laws of charge and energy-momentum. What Pauli hints at with his last remarks in brackets is the Kaluza-Klein picture, in which electric charge is interpreted as momentum in an additional space dimension in a five-dimensional spacetime.

It is not difficult to find explicit commitments from Pauli's later scientific life expressing his belief in the heuristic power of symmetry considerations. Let me just select two of them. The first is from his introduction to the "International Congress of Philosophers", held in Zürich in 1954, where Pauli states:[8]

> "It seems likely to me, that the reach of the mathematical group concept in physics is not yet fully exploited."

The second is from his closing remarks as the president of the conference "50 Years of Relativity", held in Berne in 1955, where he states with respect to

---

[7] German original (Meyenn, 1985, pp. 45–46): "Erstens scheint es mir, daß der Erhaltungssatz für Energie-Impuls dem für die Ladung doch sehr weitgehend analog ist und ich kann keinen theoretischen Grund dafür sehen, warum letzterer noch gelten sollte (wie wir es ja empirisch für den β-Zerfall wissen), wenn ersterer versagt. Zweitens müßte bei einer Verletzung des Energiesatzes auch mit dem *Gewicht* etwas sehr merkwürdiges passieren ... Dies widerstrebt meinem physikalischen Gefühl auf das äußerste! Denn es muß dann sogar auch für das Gravitationsfeld, das von dem ganzen Kasten (samt seinem radioaktiven Inhalt) selber *erzeugt* wird ... , angenommen werden, daß es sich ändern kann, während wegen der Erhaltung der Ladung das nach außen erzeugte elektrostatische Feld (beide Felder scheinen mir doch analog zu sein; das wirst Du ja übrigens auch aus Deiner fünfdimensionalen Vergangenheit noch wissen) unverändert bleiben soll."

[8] German original (Pauli, 1957b, p. 24): "Es ist mir wahrscheinlich, dass die Tragweite des mathematischen Gruppenbegriffes in der Physik heute noch nicht ausgeschöpft ist."

the still unsolved problem of whether and how the gravitational field should be described in the framework of quantum field theory:[9]

> "It seems to me, that the heart of the matter [the problem of quantizing the gravitational field] is not so much the linearity or non-linearity, but rather the presence of a more general group than the Lorentz group."

This, in fact, implicitly relates to much of the present-day research concerned with that difficult problem.

Before we can discuss specific aspects of symmetry in Pauli's work in Section 3, we now recall various aspects of symmetry principles as used in physics.

2 Remarks on the Notion of Symmetry

2.1 Spacetime

The term "symmetry" is used in such a variety of meanings, even in physics, that it seems appropriate to recall some of its main aspects. One of them is what mathematicians call an "automorphism", which basically means a "structure-preserving self-map". Take as an example (conceptually not an easy one) the modern notion of spacetime. First of all it is a set, M, the members of which are events, or better, "potential events", since we do not want to assume that every spacetime point is an actual physical event in the sense that a material happening is taking place, or at least not one which is dynamically relevant to the problem at hand.[10] That set is endowed with certain structures which are usually motivated through operational relations among actual physical events.

One such structure could be that of a preferred set of paths, which represent inertial (i.e. force free) motions of test bodies, that is, localized objects which do not react back onto spacetime structure. This defines a so-called path-structure (compare Ehlers and Koehler, 1977; Coleman and Korte, 1980), which in the simplest case reduces to an *affine structure* in which the preferred paths behave, intuitively speaking, like straight lines. This can clearly be said in a much more precise form (see, e.g., Pfister, 2004). Under very mild technical assumptions (not even involving continuity) one may then show that

---

[9] German original (Pauli, 1956, p. 267): "Es scheint mir also, daß nicht so sehr die Linearität oder Nichtlinearität der Kern der Sache ist, sondern eher der Umstand, daß hier eine allgemeinere Gruppe als die Lorentzgruppe vorhanden ist."

[10] Minkowski was well aware that empty domains of spacetime may cause conceptual problems. Therefore, in his famous 1908 Cologne address "Space and Time" (German original: "Raum und Zeit"), he said: "In order to not leave a yawning void, we wish to imagine that at every place and at every time something perceivable exists." German original (Minkowski, 1909, p. 2): "Um nirgends eine gähnende Leere zu lassen, wollen wir uns vorstellen, daß allerorten und zu jeder Zeit etwas Wahrnehmbares vorhanden ist."

the only automorphisms of that "inertial structure" can already be narrowed down to the inhomogeneous Galilei or Lorentz groups, possibly supplemented by constant scale transformations (cf. Giulini, 2006; Goldstein, 2007).[11]

Another structure to start with could have been that of a causal relation on M. This is a partial order relation which determines the pairs of points on spacetime which, in principle, could influence each other in form of a propagation process based on ordinary matter or light signals. The automorphism group of that structure is then the subgroup of bijections on M that, together with their inverse, preserve this order relation. For example, in case of Minkowski space, where the causal relation is determined by the light-cone structure, it may be shown that the most general automorphism is given by a Poincaré transformation plus a constant rescaling (Alexandrov, 1975; Zeeman, 1964). Since, according to Klein's (1872) "Erlanger Programm", any geometry may be characterized by its automorphism group, the geometry of Minkowski space is, up to constant rescalings, entirely encoded in the causal relations.

The same result can be arrived at through topological considerations. Observers (idealized to be extensionless) move in spacetime on timelike curves. Take the set $\mathcal{C}$ of all (not necessarily smooth) timelike curves which are continuous in the standard (Euclidean) topology $\mathcal{T}_E$ of Minkowski spacetime M. Now endow M with a new topology, $\mathcal{T}_P$, called the *path topology*, which is the finest topology on M which induces the same topology on each path in $\mathcal{C}$ as the standard (Euclidean) topology $\mathcal{T}_E$. The new topology $\mathcal{T}_P$ is strictly finer than $\mathcal{T}_E$ and has the following remarkable property: The automorphism group of $(M, \mathcal{T}_P)$[12], i.e. the group of bijections of M which, together with their inverses, preserve $\mathcal{T}_P$, is just the Poincaré group extended by the constant rescalings (Hawking *et al.*, 1976). This is possibly the closest operational meaning one could attribute to the topology of spacetime, since in $\mathcal{T}_P$ a set in spacetime is open if and only if every observer "times" it to be open.

All this is meant to illustrate that there are apparently different ways to endow spacetime with structures that are, physically speaking, more or less well motivated and which lead to the same automorphism group. That group may then be called the group of *spacetime symmetries*. So far, this group seems to bear no direct relation to any dynamical law. However, the physical meaning of such statements of symmetry is tied to an ontological status of spacetime points. We assumed from the onset that spacetime is a *set* M. Now, recall that Georg Cantor (1985), in his first article on transfinite set theory, started out with the following definition of a set:[13]

---

[11] We shall from now on use "Poincaré group" for "inhomogeneous Lorentz group" and "Lorentz group" for "homogeneous Lorentz group".

[12] In the standard topological way of speaking this is just the "homeomorphism group" of $(M, \mathcal{T}_P)$.

[13] German original (Cantor, 1895, p. 481): "Unter einer 'Menge' verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objecten m unserer Anschauung oder unseres Denkens (welche die 'Elemente' von M genannt werden) zu einem Ganzen."

> "By a 'set' we understand any aggregation M of definite well-distinguished objects m of our intuition or of our thinking (which are called the 'elements' of M) into a whole."

Hence we may ask: Is a point in spacetime, a "potential event" as we called it earlier, a "definite well-distinguished object of our intuition or of our thinking" ? This question is justified even though modern axiomatic set theory is more restrictive in what may be called a set (for otherwise it runs into the infamous antinomies) and also stands back from any characterization of elements in order to not confuse the axioms themselves with their possible *interpretations*.[14]

However, applications to physics require interpreted axioms, where it remains true that elements of sets are thought of as definite as in Cantor's original definition. But it is just this definiteness that seems to be physically unwarranted in application to spacetime. The modern general-relativistic viewpoint takes that into account by a quotient construction, admitting only those statements as physically meaningful that are invariant under the group of (differentiable) permutations of spacetime points. This is possible only because all other structures on spacetime, in particular the metric and with it the causal structure, are not fixed once and for all but are subsumed into the dynamical fields. Hence no non-dynamical background structures remain, except those that are inherent in the definition of a differentiable manifold. The group of automorphisms is therefore the whole diffeomorphism group of spacetime which, in some sense, comes sufficiently close to the group of all permutations.[15]

2.2 Dynamical Symmetries Versus Covariance

What is the relation between spacetime automorphisms and symmetries of dynamical laws ? Before we can answer this, we have to recall what a symmetry of a dynamical law is.

For definiteness, let us restrict our attention to dynamical laws in classical (i.e. non-quantum) physics. The equations of motion generally take the form of systems of differential equations, which we will abbreviate as **EM** (equations of motion). These equations involve two types of quantities: (1) background

---

[14] This urge for a clean distinction between the axioms and their possible interpretations is contained in the famous and amusing *dictum* attributed to David Hilbert by his student Otto Blumenthal: "One must always be able to say 'tables', 'chairs', and 'beer mugs' instead of 'points', 'lines', and 'planes'." (German original: "Man muß jederzeit an Stelle von 'Punkten', 'Geraden' und 'Ebenen' 'Tische', 'Stühle' und 'Bierseidel' sagen können.")

[15] There are clearly much more general bijections of spacetime than continuous or even differentiable ones. However, the diffeomorphism group is still n-point transitive, that is, given any two n-tuples of mutually distinct spacetime points, $(p_1, \cdots, p_n)$ and $(q_1, \cdots, q_n)$, there is a diffeomorphism ϕ such that for all $1 \leq i \leq n$ we have $\Phi(p_i) = q_i$ for all positive integers n.

structures, collectively abbreviated here by Σ, and (2) dynamical entities, collectively abbreviated here by Φ. The former will typically be represented by geometric objects on M (tensor fields, connections, etc), which are taken from a somehow specified set $\mathcal{B}$ of "admissible backgrounds". Typical background structures are external sources, like currents, and the geometry of spacetime in non-general-relativistic field theories. Dynamical entities typically involve "particles" and "fields", which in the simplest cases are represented by maps to and from spacetime,

$$\gamma : \mathbb{R} \rightarrow M \qquad (\text{“particle”})\,, \tag{1a}$$

$$\psi : M \rightarrow V \qquad (\text{“field”})\,, \tag{1b}$$

where V is usually some vector space.

In order to state the equations of motion, one has to first specify a set of so-called[16] *kinematically possible trajectories* out of which the dynamical entities Φ are taken and solutions to the equations of motion are sought. Usually this involves particle trajectories which are sufficiently smooth (typically piecewise twice continuously differentiable) and fields which are sufficiently smooth and in addition have a sufficiently rapid fall-off at large spatial distances, so as to give rise to finite quantities of energy, angular momentum, etc. This space of kinematically possible trajectories will be denoted by $\mathcal{K}$. According to the discussion above, the equation of motion takes two arguments, one from $\mathcal{B}$, the other from $\mathcal{K}$, and is hence written in the form

$$\mathbf{EM}\{\,\Sigma \mid \Phi\,\} = 0\,, \tag{2}$$

where the zero on the right-hand side may be a many-component object. Equation (2) should be read as a selection criterion on the set $\mathcal{K}$, depending on the externally specified values of Σ. We shall sometimes write $\mathbf{EM}_\Sigma$ for $\mathbf{EM}\{\,\Sigma \mid \cdot\,\}$ to denote the particular equation of motion for Φ corresponding to the choice Σ for the background structures. In general, the sets of solutions to (2) for variable Σ are Σ-dependent subset $\mathcal{D}_\Sigma \subset \mathcal{K}$, whose elements are called the *dynamically possible trajectories*[16]. We can now say more precisely what is usually meant by a symmetry:

Definition 1. An abstract group G is called a *symmetry group* of the equations of motion iff[17] the following conditions are satisfied:

1. There is an effective (see below) action $G \times \mathcal{K} \rightarrow \mathcal{K}$ of G on the set of kinematically possible trajectories, denoted by $(g, \Phi) \mapsto g \cdot \Phi$.
2. This action leaves the subset $\mathcal{D}_\Sigma \subset \mathcal{K}$ invariant; that is, for all g in G we have:

$$\mathbf{EM}\{\,\Sigma \mid \Phi\,\} = 0 \quad \Longleftrightarrow \quad \mathbf{EM}\{\,\Sigma \mid g \cdot \Phi\,\} = 0\,. \tag{3}$$

---

[16] This terminology is due to Anderson (1967).

[17] Throughout we use "iff" as abbreviation for "if and only if".

Recall that an action is called effective if no group element other than the group identity fixes all points of the set it acts on. Effectiveness is required in order to prevent mathematically trivial and physically meaningless extensions of G. What really matters are the orbits of G in $\mathcal{K}$, that is, the subsets $O_{\Phi} = \{g \cdot \Phi \mid g \in G\}$ for each $\Phi \in \mathcal{K}$. If the action were not effective, we could simply reduce G to a smaller group with an effective action and the same orbits in $\mathcal{K}$, namely the quotient group G/G', where G' is the normal subgroup of elements that fix all points of $\mathcal{K}$.

It should be noted that this definition is still very general due to the fact that no further condition is imposed on the action of G, apart from the obvious one of effectivity. For example, for fields one usually requires the action to be "local" in the sense that, for any point p of spacetime, the value $(g \cdot \psi)(p)$ of the g-transformed field should be determined by the value of the original field at some point p' of spacetime, and possibly *finitely many* derivatives of ψ at p'. If there are no dependencies on the derivatives, the action is sometimes called "ultralocal". Note that the point p' need not be identical to p, but it is assumed to be uniquely determined by g and p. A striking example of what can happen if locality is not imposed is given by the vacuum Maxwell equations (no external currents), which clearly admit the Poincaré group as ultralocally acting symmetry group. Less well known is the fact that they also admit the inhomogeneous Galilei group as symmetry group,[18] albeit the action is non-local (see Fushchich and Shtelen, 1991, or Fushchich *et al.*, 1993, Chap. 5.9). There are also other non-local symmetries of the vacuum Maxwell equations (see Fushchich and Nikitin, 1979).

The notion of symmetry is to be strictly distinguished from the notion of covariance, which we define as follows:

Definition 2. An abstract group G is called a *covariance group* of the equations of motion iff the following conditions are satisfied:

1. There is an effective action $G \times \mathcal{K} \to \mathcal{K}$ of G on the set of kinematically possible trajectories, denoted by $(g, \Phi) \mapsto g \cdot \Phi$.
2. There is also an action (this time not necessarily effective) $G \times \mathcal{B} \to \mathcal{B}$ of G on the set of background structures, likewise denoted by $(g, \Sigma) \mapsto g \cdot \Sigma$.
3. The solution-function $\Sigma \mapsto \mathcal{D}_{\Sigma} \subset \mathcal{K}$ from $\mathcal{B}$ into the subsets of $\mathcal{K}$ is G-equivariant. This means the following: If $g \cdot \mathcal{D}_{\Sigma}$ denotes the set $\{g \cdot \Phi \mid \Phi \in \mathcal{D}_{\Sigma}\}$, then, for all g in G, we have

$$g \cdot \mathcal{D}_{\Sigma} = \mathcal{D}_{g \cdot \Sigma} \,. \tag{4}$$

---

[18] This is different from, and certainly more surprising than, the better known (ultralocal) Galilei symmetry of Maxwell's equations in the presence of appropriate constitutive relations between the electric field $\boldsymbol{E}$ and the electric displacement-field $\boldsymbol{D}$ on one side, and between the magnetic induction-field $\boldsymbol{B}$ and the magnetic field $\boldsymbol{H}$ on the other; see e.g. Bellac and Lévy-Leblond (1972) and Goldin and Shtelen (2001).

> An alternative way to say this is that the relation that **EM** establishes on $\mathcal{B} \times \mathcal{K}$ via (2) is G-invariant, that is, for all g in G, we have
>
> $$\mathbf{EM}\{\,\varSigma \mid \varPhi\} = 0 \quad \Longleftrightarrow \quad \mathbf{EM}\{\,g \cdot \varSigma \mid g \cdot \varPhi\} = 0\,. \tag{5}$$

The obvious difference between (3) and (5) is that in (3) the background structure is not allowed to change. The transformed dynamical entity is required to satisfy the *very same* equation as the untransformed one, whereas for a covariance it is only required to satisfy a suitably changed set of equations. Here "changed" refers to the fact that $g \cdot \varSigma$ is generally different from $\varSigma$. Hence it is clear that a symmetry group is automatically also a covariance group, by just letting it act trivially on the set $\mathcal{B}$ of background structures. The precise partial converse is as follows: Given a covariance group G with action on $\mathcal{B}$, then for each $\varSigma \in \mathcal{B}$ define the "stabilizer subgroup" of $\varSigma$ in G as the set of elements in G that fix $\varSigma$,

$$\mathrm{Stab}_G(\varSigma) := \{g \in G \mid g \cdot \varSigma = \varSigma\}\,. \tag{6}$$

Then the subgroup $\mathrm{Stab}_G(\varSigma)$ of the covariance group is also a symmetry group of $\mathbf{EM}_\varSigma$.

The requirement of covariance is rather trivial, since it can always be met by suitably taking into account all the background structures and a sufficiently general action of G on $\mathcal{B}$. To see how this works in a specific example, consider the ordinary heat equation for the temperature field T (κ is a dimensionful constant):

$$\partial_t T - \kappa \varDelta T = 0\,. \tag{7}$$

Let $G = E_3 \times \mathbb{R}$ be the 7-parameter group of Euclidean motions (rotations and translations in $\mathbb{R}^3$) and time translations, whose defining representation on spacetime ($\mathbb{R}^3 \times \mathbb{R}$) is denoted by $g \to \rho_g$, then G acts effectively on the set of temperature fields via $g \cdot T := T \circ \rho_{g^{-1}}$ (the inverse is introduced to make this a *left* action). It is immediate from the structure of (7) that this implements G as a symmetry group of this equation. The background structures implicit in (7) are: (a) a preferred split of spacetime into space and time, (b) a preferred measure and orientation of time, and (c) a preferred distance measure on space. There are many ways to parametrize this structure, *depending on the level of generality one starts from.* If, for example, we start from special relativity, we only list those structural elements that we need *on top of* the Minkowski metric $\{\eta_{\mu\nu}\} = \mathrm{diag}(1,-1,-1,-1)$ in order to write down (7). They are given by a single constant and normalized timelike vector field n, by means of which we can write (7) in the form

$$\mathbf{EM}\{\,n \mid T\,\} := n^\mu \partial_\mu T - \kappa(n^\mu n^\nu - \eta^{\mu\nu})\partial_\mu \partial_\nu T = 0\,. \tag{8}$$

In the special class of inertial reference frames in which $n^\mu = (1,0,0,0)$, Eq. (8) reduces to (7). From the structure of (8) it is obvious that this equation admits the whole Poincaré group of special relativity as covariance group. However, the symmetry group it contains is the stabilizer subgroup of the given background structure. The latter is given by the vector field n, whose stabilizer subgroup within the Poincaré group is just $E_3 \times \mathbb{R}$, the same as for (7).

Had we started from a higher level of generality, in which no preferred coordinate systems are given to us as in special relativity, we would write the heat equation in the form

$$\mathbf{EM}\{\, n, g \mid T \,\} := n^\mu \nabla_\mu T - \kappa (n^\mu n^\nu - g^{\mu\nu}) \nabla_\mu \nabla_\nu T = 0\,, \tag{9}$$

where now n as well as g feature as background structures. n is again specified as unit timelike covariant-constant vector field, g as a flat metric, and ∇ as the unique covariant derivative operator associated to g (i.e. torsion free and preserving g). Since ∇ is here taken as a unique function of g, it does not count as independent background structure. Once again it is clear from the structure of (9) that the covariance group is now the whole diffeomorphism group of spacetime. However, the symmetry group remains the same as before since the stabilizer subgroup of the pair (g, n) is $E_3 \times \mathbb{R}$.

This example demonstrates how easy it is to almost arbitrarily inflate covariance groups by starting from higher and higher levels of generality and adding the corresponding extra structures into one's list of background structures. This possibility is neither surprising nor particularly disturbing. Slightly more disturbing is the fact that a similar game can be played with symmetries, at least at a very formal level. The basic idea is to simply declare background structures to be dynamical by letting their values be determined by equations. We may do this since we have so far not qualified equations of motion as any special sort of equations. For example, in the special relativistic context we may just take (8) and let n be determined by

$$n^\mu n^\nu \eta_{\mu\nu} = 1\,, \qquad \partial_\mu n^\nu = 0\,. \tag{10}$$

Then (8) and (10) together define a background-free (from the special relativistic point of view) system of equations for (T, n) which has the full Poincaré group as symmetry group. Its symbolic form is

$$\mathbf{EM}\{\, \emptyset \mid T, n \,\} = 0\,, \tag{11}$$

where the 0 on the right-hand side has now 18 components: one for (8), one for the first equation in (10), and 16 ($= 4 \times 4$) for the second equation in (10). But note that the T-sector of its solution space is *not* the same as that of (7), as it now also contains solutions for different n. However, as the equations (10) for n do not involve T, the total solution space for (n, T) can be thought of as fibred over the space of allowed n, with each fibre over n given by the solutions T of (8) for that given n. Each such fibre is a faithful image of the original solution space of (7), suitably transformed by a Lorentz boost that relates the original n in (7) (i.e. $\{n^\mu\} = (1, 0, 0, 0)$) to the chosen one.

Even more radically, we could take (9) and declare n and g to be dynamical entities obeying the extra equations

$$n^\mu n^\nu g_{\mu\nu} = 1\,, \qquad \nabla_\mu n^\nu = 0\,, \qquad \mathrm{Riem}[g] = 0\,, \tag{12}$$

where Riem is the Riemann curvature tensor of g, so that the last equation in (12) just expresses flatness of g. The system consisting of (9) and (12) has no background structures and admits the full diffeomorphism group as symmetry group. It is of the symbolic form

$$\mathbf{EM}\{\,\emptyset \mid T, n, g\,\} = 0\,, \tag{13}$$

which now comprises 36 components: 16 components as above and an additional set of 20 independent components of Riem. Again, note that the T-sector of the solution space of (9) is now much bigger than that of (7) or (8). With any solution T it also contains the diffeomorphism-transformed $T' = T \circ \phi^{-1}$, where $\phi \in \mathrm{Diff}(M)$. Again, since the equations for n and g do not involve T, the total solution space is fibred over the allowed n and g fields, with each fibre corresponding to a faithful image of the original solution space for (7).

Finally we remark that, in principle, constants appearing in equations of motion could also be addressed as background structures whose values might eventually be determined by more general dynamical theories. For example, one might speculate (as was done some time ago in the so-called Brans-Dicke theories) that the gravitational constant is actually the value of some field that only in the present epoch of our universe has settled to a spatially constant and quasi-static value, but whose value at much earlier times was significantly different. Another example from quantum field theory concerns the idea that masses of elementary particles are dynamically generated by the so-called Higgs field (whose existence is strongly believed but not yet experimentally confirmed).

In any case, the important message from the considerations of this subsection is the following: symmetries emerge or disappear if, respectively, background structures become dynamical ($\Sigma \to \Phi$) or dynamical structures “freeze” ($\Phi \to \Sigma$).

2.3 Observable Versus Gauge Symmetries

Within the concept of symmetry, an important distinction must be made between *observable* symmetries on one hand, and *gauge* symmetries on the other. An observable symmetry transforms a state or a history of states (trajectory) into a *different*, that is, *physically distinguishable* state or history of states. On the other hand, a gauge symmetry transforms a state or a history of states into a *physically indistinguishable* state or a history of states. In this case there is a redundancy in the mathematical description, so that the map from mathematical labels to physical states is not faithful. This is usually associated with a group, called the *group of gauge transformations*, denoted by G_{gau}, which acts on the set of state labels such that two such labels correspond to the same physical state iff they lie in the same orbit of G_{gau}.

It is clear that the notion of distinguishability introduced here refers to the set of observables, i.e. functions on state space that are physically realizable in the widest sense. Assuming for the moment that this is well defined, we could attempt a definition as follows:

> **Definition 3.** Let G be a symmetry group in the sense of Definition 1. Then $g \in G$ is called an *observable or physical symmetry* iff there exists a $\Phi \in \mathcal{D}_\Sigma$ and a physical observable that separates $g \cdot \Phi$ from Φ. If no such observable exists, g is called a *gauge symmetry.*

For a theoretician, the stipulation of what functions on state space correspond to physically realizable observables is itself hypothetical. However, it is important that *relative* to such a stipulation the distinction between observables and gauge symmetries makes sense. In the mathematical practice gauge symmetries are often signaled by an underdeterminedness of the equations of motion, which sometimes simply fail to restrict the motion in certain degrees of freedom which are then called "gauge degrees of freedom". In that case, given any solution $\Phi \in \mathcal{D}_\Sigma$, we can obtain another solution, Φ', by just changing Φ in those non-determined degrees of freedom in an arbitrary way.

For example, if the equations of motion are obtained via an action principle, such spurious degrees of freedom will typically reveal their nature through the property that motions in them are not associated with any action. As a result, the equations of motion, which are just the condition for the stationarity of the action, will not constrain the motion in these directions. Conversely, if according to the action principle the motion in some degree of freedom costs action, it can hardly be called redundant. In this sense an action principle is not merely a device for generating equations of motion, but also contains information about observables.

The combination of observable and gauge symmetries into the total symmetry group G need not at all be just a semi-direct or even direct product. Often, in field theory, the gauge group G_{gau} is indeed a subgroup of G, in fact an invariant (normal) one, but the observable symmetries, G_{obs}, are merely a quotient and not a subgroup of G. In standard group theoretic terms one says that G is a G_{gau}-extension of G_{obs}. This typically happens in electromagnetism or, more generally, in Yang-Mills type gauge theories or general relativity with globally charged configurations. In this case only the gauge transformations with sufficiently rapid fall-off at large spatial distances are proper gauge transformations in our sense, whereas the long ranging gauge transformations cost action if performed in real time[19] and therefore have to be interpreted as elements of G_{obs} (see, e.g., Giulini, 1995, and Joos *et al.*, 2003, Chap. 6).

---

[19] By the very definition of global charge, which is just the derivative of the action with respect to a long-ranging gauge transformation.

This ends our small excursion into the realm of meanings of symmetry. We now turn to the discussion of specific aspects in Pauli's work.

3 Specific Comments on Symmetries in Pauli's Work

The usage of symmetry concepts in Pauli's work is so rich and so diverse that it seems absolutely hopeless, and also inappropriate, to try to present them in a homogeneous fashion with any claim of completeness. Rather, I will comment on various subjectively selected aspects without saying that other aspects are of lesser significance. In fact, I will not include some of his most outstanding contributions, like, for example, the formulation of the exclusion principle, the neutrino hypothesis, or his anticipation of Yang-Mills gauge theory for the strong interaction.

There exist excellent reviews and discussions of these topics in the literature. Specifically I wish to refer to Bartel van der Waerden's contribution "Exclusion Principle and Spin" to the Pauli Memorial Volume (Fierz and Weisskopf, 1960, pp. 199–244), Norbert Straumann's (2004) recent lecture on the history of the exclusion principle, Pauli's own account of the history of the neutrino (Pauli, 1957a; English in Enz and Meyenn, 1994, pp. 193-217), Chien-Shiung Wu's account "The Neutrino" in the Pauli Memorial Volume (Fierz and Weisskopf, 1960, pp. 249-303), and the historical account of gauge theories by Lochlainn O'Raifeartaigh and Norbert Straumann (2000). A non-technical overview concerning "Pauli's Belief in Exact Symmetries" is given by Karl von Meyenn (1987). Last, but clearly not least, I wish to mention Charles Enz's (2002) comprehensive scientific biography of Wolfgang Pauli, which gives a detailed discussion of his scientific œvre.

In this contribution I rather wish to concentrate on some particular aspects of the notion of symmetry that are directly related to the foregoing discussion in Sections 2.2 and 2.3, as I feel that they are somewhat neglected in standard discussions of symmetry.

3.1 The Hydrogen Atom in Matrix Mechanics

In January 1926 Pauli managed to deduce the energy spectrum of the hydrogen atom from the rules of matrix mechanics. For this he implicitly used the fact that the mechanical problem of a point charge moving in a spherically symmetric force-field with a fall-off proportional to the square of the inverse distance has a symmetry group twice as large (i.e. of twice the dimension) as the group of spatial rotations alone, which it contains. Hence the total symmetry group is made of half a "kinematical" part, referring to space, and half a "dynamical" part, referring to the specific force law ($1/r^2$ fall-off). Their combination is a proper physical symmetry group that transforms physically distinguishable states into each other. In the given quantum-mechanical context one also speaks of "spectrum generating" symmetries.

Let us recall the classical problem in order to convey some idea of where the symmetries and their associated conserved quantities show up, and how they may be employed to solve the dynamical problem. Consider a mass-point of mass m and position coordinate $\boldsymbol{r}$ in the force field $\boldsymbol{F}(\boldsymbol{r}) = -(K/r^2)\boldsymbol{n}$, where r is the length of $\boldsymbol{r}$, $\boldsymbol{n} := \boldsymbol{r}/r$, and K is some dimensionful constant. Then, according to Newton's third law (an overdot stands for the time derivative),

$$\ddot{\boldsymbol{r}} = -\frac{k}{r^2}\boldsymbol{n} \qquad (k = K/m)\,. \tag{14}$$

Next to energy, there are three obvious conserved quantities corresponding to the three components of the angular-momentum vector (here written per unit mass)

$$\boldsymbol{\ell} = \boldsymbol{r} \times \dot{\boldsymbol{r}}\,. \tag{15}$$

But there are three more conserved quantities, corresponding to the components of the following vector (today called the Lenz-Runge vector),

$$\boldsymbol{e} = k^{-1}\dot{\boldsymbol{r}} \times \boldsymbol{\ell} - \boldsymbol{n}\,. \tag{16}$$

Conservation can be easily verified by differentiation of (16) using (14) and $\dot{\boldsymbol{n}} = \boldsymbol{\ell} \times \boldsymbol{n}/r^2$. Hence on has ($\ell$ = length of $\boldsymbol{\ell}$)

$$\boldsymbol{\ell} \cdot \boldsymbol{r} = 0\,, \quad \boldsymbol{\ell} \cdot \boldsymbol{e} = 0\,, \quad r + \boldsymbol{r} \cdot \boldsymbol{e} - k^{-1}\ell^2 = 0\,, \tag{17}$$

from which the classical orbit follows immediately: Setting $\boldsymbol{r} \cdot \boldsymbol{e} = re\,cos\varphi$, the last equation (17) reads

$$r = \frac{\ell^2/k}{1 + e\,\cos\varphi}\,, \tag{18}$$

which is the well-known equation for a conic section in the plane perpendicular to $\boldsymbol{\ell}$, focus at the origin, eccentricity e (= length of $\boldsymbol{e}$), and *latus rectum* $2\ell^2/k$. The vector $\boldsymbol{e}$ points from the origin to the point of closest approach (*periapsis*). The few steps leading to this conclusion illustrate the power behind the method of working with conservation laws which, in turn, rests on an effective exploitation of symmetries.

The total energy per unit mass is given by $E = \frac{1}{2}\dot{\boldsymbol{r}}^2 - k/r$. A simple calculation shows that

$$e^2 - 1 = 2E\ell^2/k^2\,, \tag{19}$$

which allows us to express E as a function of the invariants e^2 and ℓ^2. This is the relation which Pauli showed to have an appropriate matrix analog, which expresses the energy in terms of the eigenvalues of the matrices for ℓ^2 and e^2 that Pauli determined, leading straight to the Balmer formula.

From a modern point of view one would say that, for fixed energy $E < 0$,[20] the state space of this problem carries a Hamiltonian action of the Lie algebra

---

[20] For $E > 0$ one obtains a Hamiltonian action of $\mathfrak{so}(1,3)$.

$\mathfrak{so}(4)$, generated by the 3+3 quantities $\boldsymbol{\ell}$ and $\boldsymbol{e}$. Quantization then consists in the problem to represent this Lie algebra as a commutator algebra of self-adjoint operators and to determine the spectra of certain elements in the enveloping algebra. Pauli did not realize this at the time. In particular, even though he calculated the commutation relations for the six quantities $\boldsymbol{\ell}$ and $\boldsymbol{e}$, he did not realize that they formed the Lie algebra for $\mathfrak{so}(4)$, as he frankly stated much later (1955) in his address on the occasion of Hermann Weyl's 70th birthday:[21]

> "Similarly I did not know that the matrices which I had derived from the new quantum mechanics in order to calculate the energy values of the hydrogen atom were a representation of the 4-dimensional orthogonal group."

This may be seen as evidence for Pauli's superior instinct for detecting relevant mathematical structures in physics. Much later, in the CERN-report 56-31 of 1956, Pauli returned to the representation-theoretic side of this problem (published as Pauli, 1965).

3.2 Particles as Representations of Spacetime Automorphisms

The first big impact of group theory proper on physics took place in quantum theory, notably through the work of Eugene Wigner (1931) and Hermann Weyl (1928). While in atomic spectroscopy the usage of group theory could be looked upon merely as a powerful mathematical tool, it definitely acquired a more fundamental flavor in (quantum) field theory. According to a *dictum* usually attributed to Wigner, every elementary system (particle) in special-relativistic quantum theory corresponds to a unitary irreducible representation of the Poincaré group.[22] In fact, all the Poincaré invariant linear wave equations on which special-relativistic quantum theory is based, known by the names of Klein and Gordon, Weyl, Dirac, Maxwell, Proca, Rarita and Schwinger, Bargmann and Wigner, Pauli and Fierz, can be understood as projection conditions that isolate an irreducible sub-representation of the Poincaré group[23] within a reducible one that is easy to write down.

This is usually obtained as follows: Take a field ψ on spacetime M with values in a finite-dimensional complex vector space V. Let D be a finite-dimensional irreducible representation of the (double cover of the) Lorentz

---

[21] German original (Meyenn, 2001, p. 402): "Ebensowenig wußte ich, daß die Matrices, die ich ausgerechnet hatte, um die Energiewerte des Wasserstoffatoms aus der neuen Quantenmechanik abzuleiten, eine Darstellung der 4-dimensionalen orthogonalen Gruppe gewesen sind." Note that, in modern terminology, Pauli actually refers to a representation of the *Lie algebra* of the orthogonal group.

[22] The converse is not true, since there exist unitary irreducible representations (e.g. the so-called "tachyonic" representations) which cannot correspond to (real) particles.

[23] More precisely, they isolate the universal cover $\mathbb{R}^4 \rtimes \mathrm{SL}(2,\mathbb{C})$ of the Poincaré group, or sometimes an extension thereof, by the discrete transformations of space and time reversal.

group SL(2,$\mathbb{C}$) on V.[24] It is uniquely labelled by a pair (p, q) of two positive integer- or half-integer-valued numbers. In the standard terminology, $2p$ corresponds to the number of unprimed, $2q$ to the number of primed spinor indices of ψ. The set of such fields furnishes a linear representation of the (double cover of the) Poincaré group, $\mathbb{R}^4 \rtimes \mathrm{SL}(2,\mathbb{C})$, where the action of the group element $g = (a, A)$ is given by

$$g \cdot \psi := D(A)(\psi \circ g^{-1}) \,, \tag{20}$$

or for the Fourier transform $\tilde{\psi}$,

$$g \cdot \tilde{\psi} := \exp(ip_\mu a^\mu) D(A)(\tilde{\psi} \circ A^{-1}) \,. \tag{21}$$

One immediately infers from (21) that irreducibility implies that $\tilde{\psi}$ must have support on a single SL(2,$\mathbb{C}$) orbit in momentum space. Here one usually restricts oneself to those orbits consisting of non-spacelike p (those with spacelike p give rise to the tachyonic representations which are deemed unphysical), which are labeled by $p_\mu p^\mu = m^2$ with non-negative m. For ψ this means that it obeys the Klein-Gordon equation $(\Box + m^2)\psi = 0$. This is already half the way to an irreducible representation, insofar as it now contains only modes of fixed mass. But these modes still contains several spins up to the maximal value $p+q$. A second and last step then consists of projecting out one (usually the highest) spin, which gives rise to the equations named above.

In this fashion the physical meanings of *mass* and *spin* merge with the abstract mathematical meaning of mere labels of irreducible representations. Mass and spin are the most elementary attributes of physical objects, so that objects with no other attributes are therefore considered *elementary.* As just described, these elementary attributes derive from the representation theory of a group whose significance is usually taken to be that it is the automorphism group of spacetime.

However, as already discussed in Sections 1 and 2.1, this point of view presupposes a hierarchy of physical thinking in which spacetime (here Minkowski space) is considered an entity prior to (i.e. more fundamental than) matter, which may well be challenged. A more consistent but also more abstract point of view would be to think of the abstract[25] Poincaré group as prior to the matter content *as well as* the spacetime structure and to derive both simultaneously. Here "deriving" a spacetime structure (geometry) from a group would be meant in the sense of Klein's (1872) "Erlanger Programm".

---

[24] The representation D is never unitary, simply because the Lorentz group has no non-trivial finite-dimensional unitary irreducible representations. But it will give rise to an infinite-dimensional representation on the linear space of fields ψ which will indeed be unitary.

[25] "Abstract" here means to consider the isomorphicity class of the group as mathematical structure, without any interpretation in terms of transformations of an underlying set of objects.

We did already discuss in Section 1 Pauli's shift in emphasis towards a more abstract point of view as regards spacetime structure. But also regarding matter he was, next to Wigner, one of the proponents of putting symmetry considerations first and to derive the wave equations of fundamental fields as outlined above. Based on previous work by Fierz (1939) on the theory of free wave equations for higher spin, Fierz and Pauli (1939) published their very influential paper "On Relativistic Equations for Particles of Arbitrary Spin in an Electromagnetic Field" which is still much cited today.

In fact, much earlier, in his 1927 paper "Quantum Mechanics of the Magnetic Electron",[26] Pauli succeeded to implement the electron spin into non-special-relativistic quantum mechanics in an entirely representation-theoretic fashion as regards (the Lie algebra of) spatial rotations. In contrast to the other (translational) degrees of freedom, spin does not appear as the quantization of an already existent classical degree of freedom. This must have appeared particularly appealing to Pauli, who never wanted the electron spin to be understood as an intrinsic angular momentum due to a spatial rotation of a material structure. When Pauli introduced the new spin quantum number for the electron in his 1924 paper "On the Influence of the Velocity Dependence of the Electron Mass on the Zeeman Effect",[27] he deliberately stayed away from any model interpretation and cautiously referred to it as "a peculiar, classically indescribable disposition of two-valuedness of the quantum-theoretic properties of the light-electron".[28]

At that time an understandable general scepticism against possible erroneous prejudices imposed by the usage of classical models had already firmly established itself in Pauli's (and others') thinking. As much justified as this is in view of quantum mechanics, it had also led to overstatements to the effect that spin has no classical counterpart and that any classical model is even classically contradictory in the sense of violating special relativity. As regards the second point, which was also pushed by Pauli, we refer to the detailed discussion in Giulini (2007b). To the first point we first wish to mention that composite models with half-integer angular momentum states exist in ordinary quantum mechanics (without spin) as, e.g., pointed out by Bopp and Haag (1950). This is possible if their classical configuration space contains the whole group $SO(3)$ of spatial rotations. Pauli himself showed in his 1939 paper "On a Criterion for Single- or Double-Valuedness of the Eigenfunctions

---

[26] German original (Pauli, 1927): "Zur Quantenmechanik des magnetischen Elektrons".

[27] German original (Pauli, 1925): "Über den Einfluß der Geschwindigkeitsabhängigkeit der Elektronenmasse auf den Zeemaneffekt".

[28] German original (Pauli, 1925, p. 385): "... eine eigentümliche, klassisch nicht beschreibbare Art von Zweideutigkeit der quantentheoretischen Eigenschaften des Leuchtelektrons".

in Wave Mechanics"[29] the possibility of double-valued wavefunctions which give rise to half-integer angular momentum states.

Moreover, in classical mechanics there is also a precise analog of Wigner's notion of an elementary system. Recall that the space of states of a mechanical system is a symplectic manifold (phase space). The analog of an irreducible and unitary representation of the group of spacetime automorphisms is now a transitive and Hamiltonian action of this group on the symplectic manifold. It is interesting to note that this classical notion of an elementary system was only formulated much later than, and in the closest possible analogy with, the quantum mechanical one. An early reference where this is spelled out is Bacry (1967). The classification of elementary systems is now equivalent to the classification of symplectic manifolds admitting such an action (see Arens (1971 for an early reference). Here, as expected, an intrinsic angular momentum shows up as naturally as it does in quantum mechanics. What makes it slightly unusual (but by no means awkward or even inconsistent) is the fact that it corresponds to a phase space[30] that is not the cotangent bundle (space of momenta) over some configuration space of positions.

Pauli's later writings also show this strong inclination to set the fundamentals of (quantum) field theory in group-theoretic terms. In his survey "Relativistic Field Theories of Elementary Particles" (Pauli 1940), written for the 1939 Solvay Congress, Pauli immediately starts a discussion of "transformation properties of the field equations and conservation laws". His posthumously published notes on "Continuous Groups in Quantum Mechanics" (Pauli, 1965) focus exclusively on Lie-algebra methods in representation theory.

Today we are used to *define* physical quantities like energy, momentum, and angular momentum as the conserved quantities associated to spacetime automorphisms via Noether's theorem. Here, too, Pauli was definitely an early advocate of this way of thinking. Reviews on the subject written shortly after Pauli's death show clear traces of Pauli's approach (see, e.g., Kemmer *et al.*, 1959).

3.3 Spin and Statistics

Pauli's proof of the spin-statistics correlation (Pauli, 1940), first shown by Markus Fierz (1939) in his habilitation thesis, is a truly impressive example for the force of abstract symmetry principles. Here we wish to recall the basic lemmas on which it rests, which merely have to do with classical fields and representation theory.

We begin by replacing the proper orthochronous Lorentz group by its double (= universal) cover SL(2,ℂ) in order to include half-integer spin fields. We

---

[29] German original (Pauli, 1939): "Über ein Kriterium für Ein- oder Zweiwertigkeit der Eigenfunktionen in der Wellenmechanik".

[30] The phase space for classical spin is a 2-sphere, which is compact and therefore leads to a finite-dimensional Hilbert space upon quantization.

stress that everything that follows merely requires the invariance under this group. No requirements concerning invariance under space- or time-reversal are needed!

We recall from the previous section that any finite-dimensional complex representation of SL(2,$\mathbb{C}$) is labelled by an ordered pair (p,q), where p and q may assume independently all non-negative integer or half-integer values. $2p$ and $2q$ correspond to the numbers of "unprimed" and "primed" spinor indices, respectively. The tensor product of two such representations decomposes as follows:

$$D^{(p,q)} \otimes D^{(p',q')} = \bigoplus_{r=|p-p'|}^{p+p'} \quad \bigoplus_{s=|q-q'|}^{q+q'} D^{(r,s)}\,, \tag{22}$$

where – and this is the important point in what follows – the sums proceed in *integer* steps in r and s. With each $D^{(p,q)}$ let us associate a "Pauli index", given by

$$\pi : D^{(p,q)} \rightarrow ((-1)^{2p}\,,\,(-1)^{2q}) \;\in\; \mathbb{Z}_2 \times \mathbb{Z}_2\,. \tag{23}$$

This association may be extended to sums of such $D^{(p,q)}$ proceeding in integer steps, simply by assigning to the sum the Pauli index of its terms (which are all the same). Then we have[31]

$$\pi(D^{(p,q)} \otimes D^{(p',q')}) = \pi(D^{(p,q)}) \cdot \pi(D^{(p',q')})\,. \tag{24}$$

According to their representations, we can associate a Pauli index with spinors and tensors. For example, a tensor of odd/even degree has Pauli index $(-,-)/(+,+)$. The partial derivative, ∂, counts as a tensor of degree one. Now consider the most general linear (non interacting) field equations for integer spin (here and in what follows $\sum(\cdots)$ simply stands for "sum of terms of the general form $(\cdots)$"):

$$\begin{aligned} \sum \partial_{(-,-)}\Psi_{(+,+)} \;&=\; \sum \Psi_{(-,-)}\,, \\ \sum \partial_{(-,-)}\Psi_{(-,-)} \;&=\; \sum \Psi_{(+,+)}\,. \end{aligned} \tag{25}$$

These are invariant under

$$\Theta : \begin{cases} \Psi_{(+,+)}(x) \;\mapsto\; +\;\Psi_{(+,+)}(-x), \\ \Psi_{(-,-)}(x) \;\mapsto\; -\;\Psi_{(-,-)}(-x)\,. \end{cases} \tag{26}$$

Next consider any current that is a polynomial in the fields and their derivatives:

---

[31] This may be expressed by saying that the map π is a homomorphism of semigroups. One semigroup consists of direct sums of irreducible representations proceeding in integer steps with operation $\otimes$, the other is $\mathbb{Z}_2 \times \mathbb{Z}_2$, which is actually a group.

$$\begin{aligned} J_{(-,-)} = \sum \mathit{\Psi}_{(-,-)} + \mathit{\Psi}_{(+,+)}\mathit{\Psi}_{(-,-)} + \partial_{(-,-)}\mathit{\Psi}_{(+,+)} \\ +\mathit{\Psi}_{(+,+)}\partial_{(-,-)}\mathit{\Psi}_{(+,+)} + \mathit{\Psi}_{(-,-)}\partial_{(-,-)}\mathit{\Psi}_{(-,-)} + \quad \cdots \end{aligned} \tag{27}$$

Then one has

$$(\mathit{\Theta} J)(x) = -J(-x)\,. \tag{28}$$

This shows that for any solution of the field equations with charge Q for the conserved current J (Q being the space integral over J^0) there is another ($\mathit{\Theta}$-transformed) solution with charge $-Q$. It follows that charges of conserved currents cannot be sign-definite in any SL(2,$\mathbb{C}$)-invariant theory of non-interacting integer spin fields. In the same fashion one shows that conserved quantities, stemming from divergenceless symmetric tensors of rank two, bilinear in fields, cannot be sign-definite in any $\mathrm{SL}(2,\mathbb{C})$ invariant theory of non-interacting half-integer spin fields. In particular, the conserved quantity in question could be energy!

An immediate but far reaching conclusion (not explicitly drawn by Pauli) is that a relativistic generalization of Schrödinger's one-particle wave equation cannot exist. For example, for integer-spin particles, one simply cannot construct a non-negative spatial probability distribution derived from conserved four-currents. This provides a general argument for the need of second quantization, which in textbooks is usually restricted to the spin-zero case.

Upon second quantization the celebrated spin-statistics connection for free fields can now be derived in a few lines. It says that integer spin fields cannot be quantized using anti-commutators and half-integer spin fields cannot be quantized using commutators. Here the so-called Jordan-Pauli distribution plays a crucial role[32] in the (anti)commutation relations, which ensures causality (observables localized in spacelike separated regions commute). Also, the crucial hypothesis of the existence of an $\mathrm{SL}(2,\mathbb{C})$ invariant stable vacuum state is adopted. Pauli ends his paper by saying (Pauli, 1940, p. 722):

> "In conclusion we wish to state, that according to our opinion the connection between spin and statistics is one of the most important applications of the special relativity theory."

It took almost 20 years before first attempts were made to generalize this result to the physically relevant case of interacting fields by Lüders and Zumino (1958).

---

[32] The Jordan-Pauli distribution was introduced by Jordan and Pauli (1928) in their paper "Quantum Electrodynamics of Uncharged Fields" ("Zur Quantenelektrodynamik ladungsfreier Felder") in an attempt to formulate quantum electrodynamics in a manifest Poincaré invariant fashion. It is uniquely characterized (up to a constant factor) by the following requirements: (1) it must be Poincaré invariant under simultaneous transformations of both arguments; (2) it vanishes for spacelike separated arguments; (3) it satisfies the Klein-Gordon equation. The (anti)commutators of the free fields must be proportional to the Jordan-Pauli distribution, or to finitely many derivatives of it.

3.4 The Meaning of General Covariance

General covariance is usually presented as *the* characteristic feature of general relativity. The attempted meaning is that a generally covariant law takes the same form in all spacetime coordinate systems. However, in order to define the "form" of a law one needs to make precisely the distinction between background entities, which are constitutive elements of the law, and the dynamical quantities which are to obey the laws so defined (cf. Section 2.1). In the language we introduced above, general covariance cannot just mean simple covariance under all smooth and invertible transformations of spacetime points (i.e., that the spacetime diffeomorphism group is a covariance group is the sense of Definition 2), for that would be easily achievable without putting any restriction on the intended law proper, as was already pointed out by Erich Kretschmann (1917). Einstein agreed with Kretschmann's criticism, which he called "acute" (German original: "scharfsinnig"; Einstein, 1918), and withdrew from the view that the principle of general covariance has at least some heuristic power in the following sense:[33]

> "Between two theoretical systems which are compatible with experience, that one is to be preferred which is the simpler and more transparent one from the standpoint of absolute differential calculus. Try to bring Newton's gravitational mechanics in the form of generally covariant equations (four-dimensional) and one will surely be convinced that principle a)[34] is, if not theoretically, but practically excluded."

But the *principle of general covariance* is intended as a non-trivial selection criterion. Hence modern writers often characterize it as the requirement of diffeomorphism invariance, i.e. that the diffeomorphism group of spacetime is a symmetry group in the sense of Definition 1. But then, as we have seen above, the principle is open to trivializations if one allows background structures to become formally dynamical. This possibility can only be inhibited if one limits the amount of structure that may be added to the dynamical fields.[35]

---

[33] German original (Einstein, 1918, p. 242): "Von zwei mit der Erfahrung vereinbaren theoretischen Systemen wird dasjenige zu bevorzugen sein, welches vom Standpunkte des absoluten Differentialkalküls das einfachere und durchsichtigere ist. Man bringe einmal die Newtonsche Gravitationsmechanik in die Form von kovarianten Gleichungen (vierdimensional) und man wird sicherlich überzeugt sein, daß das Prinzip a) diese Theorie zwar nicht theoretisch, aber praktisch ausschließt."

[34] Einstein (1918, p. 241) formulates principle a) thus: "Principle of relativity: The laws of nature exclusively contain statements about spacetime coincidences; therefore they find their natural expression in generally covariant equations."

[35] Physically speaking, one may be tempted to just disallow such formal 'equations of motion' whose solution space is (up to gauge equivalence) zero dimensional. But this would mean that one would have to first understand the solution space of a given theory before one can decide on its general covariance properties, which would presumably render it a practically fairly useless criterion.

The reason why I mention all this here is that Pauli's relativity article is, to my knowledge, the only place in the literature that addresses this point, albeit not as explicitly as one might wish. After mentioning Kretschmann's objection, Pauli (1958b, p. 150) remarks (Pauli's emphases): [36]

> "The generally covariant formulation of the physical laws acquires a physical content only through the principle of equivalence, in consequence of which gravitation is described *solely* by the g_{ik} and the latter are not given independently from matter, but are themselves determined by the field equations. Only for this reason can the g_{ik} be described as *physical quantities*."

Note how perceptive Pauli addresses the two central issues: 1) that one has to limit the the amount of dynamical variables and 2) that dynamical structures have to legitimate themselves as physical quantities through their back reaction onto other (matter) structures. It is by far the best few-line account of the issue that I know of, though perhaps a little hard to understand without the more detailed discussion given above in Section 2.2. Most modern textbooks do not even address the problem. See Giulini (2007a) for more discussion.

3.5 General Covariance and Antimatter

In this section I wish to give a brief but illustrative example from Pauli's work for the non-trivial distinction between observable physical symmetries on one hand and gauge symmetries on the other (cf. Section 2.3). The example I have chosen concerns an argument within the (now outdated) attempts to understand elementary particles as regular solutions of classical field equations. Pauli reviewed such attempts in a rather detailed fashion in his relativity article, with particular emphasis on Weyl's theory, to which he had actively contributed in two of his first three published papers in 1919.

The argument proper says that in any generally covariant[37] theory, which allows for regular static solutions representing charged particles, there exists for any solution with mass m and charge e another such solution with the same mass but opposite charge $-e$. Pauli's proof looks like an almost trivial application of general covariance and runs as follows: Let $g_{\mu\nu}(x^\lambda)$ and $A_\mu(x^\lambda)$ represent the gravitational and electromagnetic field, respectively. The hypothesis of staticity implies that coordinates (and gauges for A_μ) can be chosen such that all fields are independent of the time coordinate, x^0, and

---

[36] German original (Pauli, 2000b, p. 181): "Einen physikalischen Inhalt bekommt die allgemein kovariante Formulierung der Naturgesetze erst durch das Äquivalenzprinzip, welches zur Folge hat, daß die Gravitation durch die g_{ik} *allein* beschrieben wird, und daß diese nicht unabhängig von der Materie gegeben, sondern selbst durch Feldgleichungen bestimmt sind. Erst deshalb können die g_{ik} als *physikalische Zustandsgrößen* bezeichnet werden."

[37] Here "general covariance" is taken to mean that the diffeomorphism group of spacetime acts as symmetry group.

that $g_{0i} \equiv 0$ as well as $A_i \equiv 0$ for $i = 1, 2, 3$.[38] Now consider the orientation-reversing diffeomorphism $\phi : (x^0, \boldsymbol{x}) \mapsto (-x^0, \boldsymbol{x})$. It maps the gravitational field to itself while reversing the sign of A_0 and, hence, of the electric field. General covariance assures these new fields to be again solutions with the same total mass but opposite total electric charge.

Pauli presents this argument in his second paper addressing Weyl's theory, entitled "To the Theory of Gravitation and Electricity by Hermann Weyl"[39] and also towards the end of Section 67 of his relativity article. The idea of this proof is due to Weyl who communicated it (without formulae) in his first two letters to Pauli (Doc. [1] and [2] in Hermann *et al.*, 1979) as Pauli also acknowledges in his paper (Pauli, 1919b, p. 462, footnote 2).

It is interesting to note that Einstein (1925) rediscovered the very same argument and found it worthy of a separate communication. At the time it was common to all, Weyl, Pauli, and Einstein, to regard the argument a nuisance and of essentially destructive nature. This was because at this time antiparticles had not yet been discovered so that the apparent asymmetry as regards the sign of the electric charges of fundamental particles was believed to be a fundamental property of nature. Already in his first paper on Weyl's theory, entitled "Perihelion Motion of Mercury and Deflection of Rays in Weyl's Theory of Gravitation"[40], Pauli emphasized:[41]

> "The main difficulty [with Weyl's theory] is – apart from Einstein's objection, which appears to me not yet sufficiently disproved – that the theory cannot account for the asymmetry between the two sorts of electricity."

Now, there is an interesting conceptual point hidden in this argument that relates to our discussions in Sections 2.2 and 2.3. First of all, the two solutions are clearly considered physically distinct, otherwise the argument could not be understood as contradicting the charge asymmetry in nature. Hence the diffeomorphism involved cannot be considered a gauge transformation but rather corresponds to a proper physical symmetry. On the other hand, we know that diffeomorphisms within bounded regions must be considered as gauge transformations, for otherwise one would run into the dilemma set by the so-called "hole argument"[42]. Hence one faces the problem of how to

---

[38] The latter conditions distinguish staticity from mere stationarity. The condition on A_i may, in fact, be relaxed.

[39] German original (Pauli, 1919b): "Zur Theorie der Gravitation und der Elektrizität von Hermann Weyl".

[40] German original (Pauli, 1919a): "Merkurperihelbewegung und Strahlenablenkung in Weyls Gravitationstheorie".

[41] German original (Pauli, 1919a, p. 749): "Die Hauptschwierigkeit ist – neben Einstein's Einwand, der mir durchaus noch nicht hinreichend widerlegt scheint –, daß die Theorie von der Asymmetrie der beiden Elektrizitätsarten nicht befriedigend Rechenschaft zu geben vermag."

[42] Let Ω be a bounded region in spacetime which is disjoint from a spacelike hypersurface Σ. Consider two solutions to the field equations which merely differ by the

characterize those diffeomorphisms which are not to be considered as gauge transformations (cf. Section 2.3).

It is conceivable that this question is not decidable without contextual information; see, e.g., Giulini (1995) and Joos *et al.* (2003, Chap. 6) for more discussion of this point. The historical sources have almost nothing to say about this, though there are suggestions by all three mentioned authors how to circumvent the argument by adding more non-dynamical structures, as a result of which general covariance is lost. Einstein, being most explicit here, suggested the existence of a timelike vector field which fixes a time orientation. At least the so-defined time orientation would then have to be considered as a non-dynamical structure of type Σ (cf. Section 2.2) in order to break the symmetry group down to the stabilizer group of Σ. The time-orientation-reversing transformation used above would then not be a symmetry anymore. Similar suggestions were made by Weyl, who also hinted at a structure to distinguish past and future:[43]

> "Their essential difference [of past and future] I take, contrary to most physicists, to be a fact of much more fundamental meaning than the essential difference between positive and negative charge."

In the last (fifth) edition of "Raum Zeit Materie", Weyl writes regarding his unified theory (Weyl's emphases):[44]

---

action of a diffeomorphism ϕ with support in Ω. If they are considered distinct, then the theory cannot have a well posed initial-value problem, since then for any Σ distinct solutions exist with identical data on Σ. This is a rephrasing of Einstein's original argument (Einstein, 1914a, p. 178; Einstein and Grossmann, 1914a, p. 260; Einstein and Grossmann, 1914b, p. 218), which did not construct a contradiction to the existence of a well posed *initial-value problem*, but rather to the requirement that the gravitational field be determined by the matter content (more precisely: its energy momentum tensor). But this requirement is clearly never fulfilled in any generally covariant theory in which the gravitational field has its own degrees of freedom, independent of whether one regards diffeomorphisms as gauge. Slightly later Einstein rephrased it so as to construct a contradiction to the existence of a well-posed *boundary-value problem* (Einstein, 1914b, p. 1167), which is also not the right thing to require from equations that describe the propagation of fields with their own degrees of freedom.

[43] German original by Weyl (Hermann *et al.*, 1979, p. 6): "Ihren Wesensunterschied [von Vergangenheit und Zukunft] halte ich, im Gegensatz zu den meisten Physikern, für eine Tatsache von noch viel fundamentalerer Bedeutung als der Wesensunterschied zwischen positiver und negativer Elektrizität."

[44] German original (Weyl, 1991, p. 308): "Die Theorie gibt keinen Aufschluß über die *Ungleichartigkeit von positiver und negativer Elektrizität.* Das kann ihr aber nicht zum Vorwurf gemacht werden. Denn jene Ungleichartigkeit beruht ohne Zweifel darauf, daß von den beiden Urbestandteilender der Materie, Elektron und Wasserstoffkern, der positiv geladene mit einer anderen Masse verbunden ist als der negaiv geladene; sie entspringt aus der Natur der Materie und nicht des Feldes."

> "The theory gives no clue as regards the *disparity of positive and negative electricity.* But that cannot be taken as a reproach against the theory. For that disparity is based without doubt on the fact that of both fundamental constituents of matter, the electron and the hydrogen nucleus, the positively charged one is tight to another mass then the negatively charged one; it originates from the nature of matter and not of the field."

Given that Weyl is talking about his unified field theory of gravity and electricity, whose original claim was to explain all of matter by means of field theory, this statement seems rather surprising. It may be taken as a sign of Weyl's beginning retreat from his once so ambitious program.

3.6 Missed Opportunities

3.6.1 Supersymmetry

One issue that attracted much attention during the 1960s was whether the observed particle multiplets could be understood on the basis of an all-embracing symmetry principle combining the Poincaré group with the internal symmetry groups displayed by the multiplet structures. This combination of groups should be non-trivial, i.e., not be a direct product, for otherwise the internal symmetries would commute with the spacetime symmetries and lead to multiplets degenerate in mass and spin (see, e.g. Raifeartaigh, 1965). Subsequently, a number of no-go theorems appeared, which culminated in the now most famous theorem of Coleman and Mandula (1967). This theorem states that those generators of symmetries of the S-matrix belonging to the Poincaré group necessarily commute with those belonging to internal symmetries. The theorem is based on a series of assumptions[45] involving the crucial technical condition that the S-matrix depends analytically on standard scattering parameters. What is less visible here is that the structure of the Poincaré group enters in a decisive way. This result would not follow for the Galilei group, as was explicitly pointed out by Coleman and Mandula (1967, p. 159).

One way to avoid the theorem of Coleman and Mandula is to generalize the notion of symmetries. An early attempt was made by Golfand and Likhtman (1971), who constructed what is now known as a super Lie algebra, which generalizes the concept of a Lie algebra (i.e., symmetry generators obeying certain commutation relations) by also involving anti-commutators. In this way it became possible for the first time to link particles of integer and half-integer spin by a symmetry principle. It is true that supersymmetry

---

[45] The assumptions are: (1) there exists a non-trivial (i.e., $\neq \mathbf{1}$) S-matrix which depends analytically on s (the squared centre-of-mass energy) and t (the squared momentum transfer); (2) the mass spectrum of one-particle states consists of (possibly infinite) isolated points with only finite degeneracies; (3) the generators (of the Lie algebra) of symmetries of the S-matrix contain (as a Lie sub-algebra) the Poincaré generators; (4) some technical assumptions concerning the possibility of writing the symmetry generators as integral operators in momentum space.

still maintains the degeneracy in masses and hence cannot account for the mass differences in multiplets. But its most convincing property, the symmetry between bosons and fermions, suggested a most elegant resolution of the notorious ultraviolet divergences that beset quantum field theory.

It is remarkable that the idea of a cancelation of bosonic and fermionic contributions to the vacuum energy density occurred to Pauli. In his lectures "Selected Topics in Field Quantization", delivered in 1950–1951, he posed the question of (Pauli, 2000a, p. 33)

> "whether these zero-point energies [from bosons and fermions] can compensate each other."

He tried to answer this question by writing down the formal expression for the zero-point energy density of a quantum field of spin j and mass $m_j > 0$ (Pauli restricted his attention to spin 0 and spin 1/2, but the generalization is immediate):

$$4\pi^2 \frac{E_j}{V} = (-1)^{2j}(2j+1) \int dk\, k^2 \sqrt{k^2+m^2}\,. \tag{29}$$

Cancelation should take place for high values of k. The expansion

$$4\int_0^K dk\, k^2\sqrt{k^2+m^2} = K^4 + m_j^2 K^2 - m_j^4 \log(2K/m_j) + O(K^{-1}) \tag{30}$$

shows that the quartic, quadratic, and logarithmic terms must cancel in the sum over j for the limit $K \to \infty$ to exist. This implies that for $n = 0, 2, 4$ one must have

$$\sum_j (-1)^{2j}(2j+1)m_j^n = 0 \quad \text{and} \quad \sum_j (-1)^{2j}(2j+1)\log(m_j) = 0\,. \tag{31}$$

Pauli (2000a, p. 33) comments that

> "these requirements are so extensive that it is rather improbable that they are satisfied in reality."

Unless enforced by an underlying symmetry, one is tempted to add! This would have been the first call for a supersymmetry in the year 1951.

However, the real world does not seem to be as simple as that. Supersymmetry, if at all existent, is strongly broken in the phase we live in. So far no supersymmetric partner of any existing particle has been detected, even though some of them (e.g., the neutralino) are currently suggested to be viable candidates for the missing-mass problem in cosmology. Future findings (or non-findings) at the Large Hadron Collider (LHC) at CERN will probably have a decisive impact on the future of the idea of supersymmetry which – whether or not it is realized in nature – is certainly very attractive; and Pauli came close to it.

3.6.2 Kaluza-Klein Monopoles

Ever since its first formulation in 1921, Pauli as well as Einstein were much attracted by a geometric idea of Theodor Kaluza and its refinement by Oskar Klein. According to this idea the classical theories of the gravitational and the electromagnetic field could be unified into a single theory, in which the unified field has the same meaning as Einstein's gravitational field in general relativity, namely as a metric tensor of spacetime, but now in five instead of four dimensions. The momentum of a particle in the additional fifth direction (which is spacelike) is now to be interpreted as its charge. Charge is conserved because the geometry of spacetime is *a priori* restricted to be independent of that fifth direction. The combined field equations are exactly the five-dimensional analog of Einstein's equations for general relativity.

A natural question to address in this unified classical theory was whether it admits solutions that could represent particle-like objects. More precisely, the solution should be stationary, everywhere regular, and possess long-ranging gravitational and electromagnetic fields (usually associated with aspects of mass and charge). Pauli, who was very well familiar with this theory since its first appearance,[46] kept an active interest in it even after the formulations of quantum mechanics and early quantum electrodynamics. It made it unquestionable for him that the problem of matter could not be adequately addressed in the framework of a classical field theory, unlike Einstein, who maintained such a hope in various forms until the end of his life in 1955.

It is therefore remarkable that in 1943 Einstein and Pauli wrote a paper in which they proved the non-existence of such solutions. The introduction contains the following statement (Einstein and Pauli, 1943, p. 131):

> "When one tries to find a unified theory of the gravitational and electromagnetic fields, he cannot help feeling that there is some truth in *Kaluza's* five-dimensional theory."

In fact, Einstein and Pauli offered a proof for the more general situation with an arbitrary number of additional space dimensions, fulfilling the generalized Kaluza-Klein "cylinder condition" that the gravitational field should not depend on any of these extra directions. Note that this extra condition introduces non-dynamical background structures, so that of the five-dimensional diffeomorphism group only those diffeomorphisms preserving this condition can act as symmetries, a point Pauli often emphasized as a deficiency regarding the Kaluza-Klein approach.

Restricting attention to five dimensions, the explicitly stated hypotheses underlying the proof were these (Einstein and Pauli, 1943, p. 131; annotations in square brackets within quotations are mine):

---

[46] It came out too late to be considered in the first edition of Pauli's relativity article. But he devoted comparatively large space to it in his supplementary notes written in early 1956 for the first English edition (Pauli, 1958b, suppl. note 23, pp. 227-232; Pauli, 2000b, pp. 276-282)

- H1 “The field is stationary (i.e the g_{ik} [the five-dimensional metric] are independent of x^4 [the time coordinate]).” Clearly, g_{ik} is also assumed to be independent of the fifth coordinate x^5.
- H2 “It [the field g_{ik}] is free from singularities.”
- H3 “It is imbedded in a Euclidean space (of the Minkowski type), and for large values of r (r being the distance from the origin of the spatial coordinate system) g_{44} has the asymptotic form $g_{44} = -1 + \mu/r$, where $\mu \neq 0$.” This condition is meant to assure the non-triviality of the solution, i.e. that there really is an attracting object at the spatial origin. This becomes clear if one recalls that in the lowest weak-field and slow-motion approximation $1 + g_{44}$ just corresponds to the Newtonian gravitational potential. Unfortunately, the other statement: “It is imbedded in a Euclidean space (of the Minkowski type)” seems ambiguous, since the solution is clearly not meant to be just a (portion of) five-dimensional flat Minkowski space. Hence the next closest reading is presumably that the underlying five-dimensional spacetime manifold is (diffeomorphic to) $\mathbb{R}^5$, with some non-flat metric of Minkowskian signature $(-,+,+,+,+)$.[47]

The elegant method of the proof makes essential use of the fact that the suitably restricted group of spacetime diffeomorphisms (to those preserving the cylinder condition) is a symmetry group for the full set of equations in the sense of (3) of Definition 1. More precisely, two types of diffeomorphisms from that class are considered separately by Einstein and Pauli:

- D1 arbitrary diffeomorphisms in the three coordinates (x^1, x^2, x^3) which leave invariant the (x^4, x^5) coordinates;
- D2 linear diffeomorphisms in the (x^4, x^5) coordinates, leaving invariant the (x^1, x^2, x^3) coordinates.

Now, as a matter of fact, this innocent looking split introduces a further and, as it turns out, crucial restriction, over and above the hypotheses H1–H3. The point is that the split and, in particular, the set D2 of diffeomorphisms simply do not exist unless the spacetime manifold, which in H3 was assumed to be $\mathbb{R}^5$, globally splits into $\mathbb{R}^2 \times \mathbb{R}^3$ such that the first factor, $\mathbb{R}^2$, corresponds to the x^4x^5-planes of constant spatial coordinates (x^1, x^2, x^3) and the second factor, $\mathbb{R}^3$, corresponds to the $x^1x^2x^3$-spaces of constant coordinates (x^4, x^5). But this need not be the case if H1–H3 are assumed. The identity derived by Einstein and Pauli from the requirement that transformations of the field induced by diffeomorphisms of the type D2 are symmetries are absolutely crucial in proving the non-existence of regular solutions.[48]

---

[47] In fact, it turns out that formally the proof does not depend on whether the fifth dimension is space- or time-like, as noted by Einstein and Pauli (1943, p. 134).

[48] Specifically we mean their identity (13), which together with spatial regularity implies the integral form (13a), which in turn leads directly to vanishing mass in (22–23a). (All references are to their formulae in Einstein and Pauli, 1943.)

We now know that this additional restriction is essential to the non-existence result: There do exist solutions of the type envisaged that satisfy H1–H3, but violate the extra (and superfluous) splitting condition.[49] They are called *Kaluza-Klein monopoles* (Sorkin, 1983; Gross and Perry, 1983) and carry a gravitational mass as well as a magnetic charge. It is hard to believe that Pauli as well as Einstein would not have been much impressed by those solutions, though possibly with different conclusions, had they ever learned about them. It is also conceivable that these solutions could have been found at the time, had real attempts been made, rather than – possibly – discouraged by Pauli's and Einstein's result. In fact, Kurt Gödel, who was already in Princeton when Pauli visited Einstein, found his famous cosmological solution (Gödel, 1949) by a very similar geometric insight that also first led to the Kaluza-Klein monopoles (Sorkin, 1983).[50]

3.7 Irritations and Psychological Prejudices

One of Pauli's major interests were discrete symmetries, in particular the transformation of space inversion, $\boldsymbol{x} \mapsto -\boldsymbol{x}$, also called *parity* transformation. Given a linear wave equation which is symmetric under the proper orthochronous (i.e. including no space and time inversions) Poincaré group, one may ask whether it is also symmetric under space and time inversions. For this to be a well defined question one has to formulate conditions on how these inversions interact with Poincaré transformations. Let us focus on the operation of space inversion. If this operation is implementable by an operator $\mathcal{P}$, it must conjugate each rotation and each time translation to their respective self, and each boost and each space translation to their respective inverse. This follows simply from the geometric meaning of space inversion. Hence, generally speaking, we need to distinguish the following three possible scenarios (recall the notation from Section 2.2):

(a) $\mathcal{P}$ acts on $\mathcal{K}$ and is a symmetry, i.e. leaves $\mathcal{D}_\Sigma \subset \mathcal{K}$ invariant;
(b) $\mathcal{P}$ acts on $\mathcal{K}$ and is no symmetry, i.e. leaves $\mathcal{D}_\Sigma \subset \mathcal{K}$ not invariant;
(c) $\mathcal{P}$ is not implementable on $\mathcal{K}$.

---

[49] The somewhat intricate topology of the Kaluza-Klein spacetime is this: The x^5 coordinate parametrizes circles which combine with the 2-spheres (polar coordinates (θ, φ)) of constant spatial radius, r, into 3-spheres (Hopf fibration) which are parametrized by (θ, φ, x^5), now thought of as Euler angles. The radii of these 3-spheres appropriately shrink to zero as r tends to zero, so that $(r, \theta, \varphi, x^5)$ define, in fact, polar coordinates of $\mathbb{R}^4$. Together with time, x^4, we get $\mathbb{R}^5$ as global topology. Now the submanifolds of constant (x^1, x^2, x^3) are those of constant (r, θ, φ) and have a topology $\mathbb{R} \times S^1$ rather than $\mathbb{R}^2$, so that the linear transformations D2 in the (x^4, x^5) coordinates do not define diffeomorphisms of the Kaluza-Klein spacetime manifold.

[50] Both use invariant metrics on 3-dimensional group manifolds, $SU(2)$ in the Kaluza-Klein case, $SU(1,1)$ in Gödel's case. This simplifies the calculations considerably.

It is clear that when one states that a certain equation is not symmetric under $\mathcal{P}$ one usually addresses situation (b), though situation (c) also occurs, as we shall see.

Consider now the field of a massless spin-$\frac{1}{2}$ particle that transforms irreducibly under the proper orthochronous Poincaré group. The field is then either a two-component spinor, ϕ^A, which in the absence of interactions obeys the so-called Weyl equation[51]

$$\partial_{AA'}\phi^A = 0\,. \tag{32}$$

Alternatively, one may also start from a four-component Dirac spinor,

$$\psi = \begin{pmatrix} \phi^A \\ \overline{\chi}_{A'} \end{pmatrix} \tag{33}$$

which carries a reducible representation of the proper orthochronous Poincaré group: If ϕ^A transforms with $A \in SL(2,\mathbb{C})$ then $\overline{\chi}_{A'}$ transforms with $(A^\dagger)^{-1}$ (being an element of the complex-conjugate dual space), so that the space of the upper two components ϕ^A of ψ and the space of the lower two components $\overline{\chi}_{A'}$ of ψ are separately invariant. One may then eliminate two of the four components by the so-called Majorana condition, which requires the state ψ to be identical with its charge-conjugate, ψ^c, where

$$\mathcal{C} : \psi \mapsto \psi^c := i\gamma^2\psi^* = \begin{pmatrix} \chi^A \\ \overline{\phi}_{A'} \end{pmatrix}\,. \tag{34}$$

Hence for a Majorana spinor one has $\phi = \chi$ and the interaction-free Dirac equation reads

$$\gamma^\mu\partial_\mu\psi := \sqrt{2}\begin{pmatrix} 0 & \partial^{AA'} \\ \partial_{A'A} & 0 \end{pmatrix}\begin{pmatrix} \phi^A \\ \overline{\phi}_{A'} \end{pmatrix} = 0\,. \tag{35}$$

One can now either regard (32) or (35) as the interaction-free equation for a neutrino.

Here I wish to briefly recall a curious discussion between Pauli and Fierz on whether or not these two equations describe physically different states of affairs. Superficially this discussion is about a formal and, mathematically speaking, rather trivial point. But, as we will see, it relates to deep-lying preconceptions in Pauli's thinking about issues of symmetry. This makes it worth looking at this episode in some detail.

---

[51] Here I use the standard spinor notation where upper-case capital Latin indices refer to (components of) elements in spinor space (2-dimensional complex vector space), lower case indices to the dual space, and primed indices to the respective complex-conjugate spaces. Indices are raised and lowered by using a (unique up to scale) SL(2,$\mathbb{C}$) invariant 2-form. An overbar denotes the map into the complex-conjugate vector space. Unless stated otherwise, my conventions are those of Sexl and Urbantke (2001).

First note that there is an obvious bijection, β, between two-component spinors and Majorana spinors, given by

$$\beta : \phi^A \mapsto \begin{pmatrix} \phi^A \\ \overline{\phi}_{A'} \end{pmatrix} . \tag{36}$$

Note also that the set of Majorana spinors is *a priori* a *real*[52] vector space, though it has a complex structure, j, given by

$$j : \begin{pmatrix} \phi^A \\ \overline{\phi}_{A'} \end{pmatrix} \mapsto \begin{pmatrix} i\phi^A \\ -i\overline{\phi}_{A'} \end{pmatrix} , \tag{37}$$

with respect to which the bijection (36) satisfies $\beta \circ i = j \circ \beta$, where i stands for the standard complex structure (multiplication with imaginary unit i) in the space $\mathbb{C}$ of two-component spinors. However, regarded as a map between complex vector spaces, the bijection β is *not* linear.

Now, Pauli observed already in 1933 that the Weyl equation (32) is not symmetric under parity. Hence he concluded that it could not be used to describe nature. In fact, parity cannot even be implemented as a *linear* map on the space of two-component spinors (case (c) above). This is easy to see and in fact true for any irreducible representation of the Lorentz group that stays irreducible if restricted to the rotation group (i.e. for purely primed or purely unprimed spinors).[53]

On the other hand, the Dirac equation *is* symmetric under space inversions. Indeed, the spinor-map corresponding to the inversion in the spatial plane perpendicular to the timelike normal n is given by

$$\mathcal{P} : \psi \mapsto \psi^p := \eta\, n_\mu \gamma^\mu (\psi \circ \rho_n) , \tag{38}$$

where $\rho_n : x^\mu \mapsto -x^\mu + 2n^\mu(n_\nu x^\nu)$, and η is a complex number of unit modulus, called the *intrinsic parity* of the particular field ψ. It is easy to see that $\mathcal{P}$ is a symmetry of (35) for any η. Note that $\mathcal{P}^2 = \eta^2\mathbf{1}$ so that $\eta \in \{1, -1, i, -i\}$, since for spinors one only requires $\mathcal{P}^2 = \pm\mathbf{1}$ (rather than $\mathcal{P}^2 = \mathbf{1}$). It is also easy to verify that $\mathcal{P}$ commutes with $\mathcal{C}$ iff $\eta = \pm i$. So if we assign imaginary parity to the Majorana field[54], the operator $\mathcal{P}$ also acts

---

[52] The reality structure on the complex vector space of Dirac spinors is provided by the charge conjugation map.

[53] As stated above, the geometric meaning of space inversion requires that the parity operator (if existent) commutes with spatial rotations and conjugates boosts to their inverse. The first requirement implies (via Schur's Lemma) that it must be a multiple of the identity in any irreducible representation that stays irreducible when restricted to the rotation subgroup, which contradicts the second requirement. Hence it cannot exist in such representations, which are precisely those with only unprimed or only primed indices.

[54] This is also the standard choice in quantum field theory; see e.g. Weinberg (1995, pp. 126, 226).

on the subspace of Majorana spinors. We conclude that the free Majorana equation *is* parity invariant.

Hence it seems at first that the Weyl formulation and the Majorana formulation differ since they have different symmetry properties. But this is not true. Using the bijection (36), we can pull back the parity map (38) to the space of two-component spinors, where it becomes (now either $\eta = i$ or $\eta = -i$)

$$\phi^A \mapsto \eta\sqrt{2}\, n^{AA'}(\overline{\phi}_{A'} \circ \rho_n)\,, \tag{39}$$

which is now an anti-linear map on the space of two-component spinors.

All this was essentially pointed out to Pauli by Fierz in a letter dated February 6, 1957, (Meyenn, 2005a, p. 171) in connection with Lee's and Yang's two-component theory of the neutrino. Fierz correctly concluded from this essential equivalence[55] that the 2-component theory as such (i.e. without interactions) did not warrant the conclusion of parity violation; only interactions could be held responsible for that.

This was a relevant point in the theoretical discussion at the time, as can be seen from the fact that there were two independent papers published in *The Physical Review* shortly after Fierz's private letter to Pauli, containing the very same observation. The first paper was submitted on February 13 by McLennan (1957), the second on March 25 by Case (1957). In fact, Serpe (1952) made this observation already in 1952 and emphasized it once more in 1957 (Serpe, 1957).

One might be worried about the anti-linearity of the transformation in (39). In that respect, also following Fierz, an illuminating analogy may be mentioned regarding the vacuum Maxwell equations, which can be written in the form

$$i\partial_t\boldsymbol{\Phi} - \boldsymbol{\nabla} \times \boldsymbol{\Phi} = 0\,, \qquad \boldsymbol{\nabla} \cdot \boldsymbol{\Phi} = 0\,, \tag{40}$$

where

$$\boldsymbol{\Phi} := \boldsymbol{E} + i\boldsymbol{B} \tag{41}$$

is a complex combination of the electric and magnetic field. Both equations (40) are clearly equivalent with the full set of Maxwell's equations. It can be shown that spatial inversions cannot be implemented as complex-linear transformations on the complex-valued field $\boldsymbol{\Phi}$.[56] But, clearly, we know that Maxwell's equations are parity invariant, namely if we transform the electric field as $\boldsymbol{E} \mapsto -\boldsymbol{E} \circ \rho$ (polar vector-field) and the magnetic field as $\boldsymbol{B} \mapsto \boldsymbol{B} \circ \rho$ (axial vector-field), where $\rho : (t, \boldsymbol{x}) \mapsto (t, -\boldsymbol{x})$. This corresponds to an *anti*linear symmetry of (40), given by $\boldsymbol{\Phi} \mapsto -\overline{\boldsymbol{\Phi}} \circ \rho$.

---

[55] More precisely, this equivalence means the existence of a bijection that maps all quantities of interest (states, currents, symmetries) of one theory to the other.

[56] Equations (40) are equivalent to $\partial^{AA'} f_{AB} = 0$, where f_{AB} is the unprimed spinor equivalent of the tensor $F_{\mu\nu}$ for the electromagnetic field strength. Parity cannot be linearly implemented on this purely unprimed spinor, for reasons already explained in footnote 53.

Coming back to Fierz's (and others') original observation for the spinor field, they were accepted without much ado by others. For example, in her survey on the neutrino in the Pauli Memorial Volume, Wu states: "It is the interaction and the interaction only that violates parity" (Fierz and Weisskopf, 1960, footnote p. 270). In note 25c of that paper she explicitly thanks Fierz for "enlightening discussions" on the two-component theory of the neutrino. Clearly Fierz expected his observation to be of interest to Pauli, who had already in the 1933 first edition of his handbook article on wave mechanics propagated the view that Weyl's two-component equations are[57]

> "not invariant under reflections (interchange of left and right) and, as a consequence, not applicable to the physical reality."

But instead, Pauli reacts with a surprising plethora of ridiculing remarks:[58]

> "Dear Mr. Fierz! Your letter from the 6th is the biggest blunder you ever commited in your life! (Probably this afternoon you will send a correction). Have only read the first paragraph of your letter which originated in the asylum and was shaking with laughter. ... When this letter arrives (yours I will frame!) you probably will already know everything."

Personal irritations emerged which lasted about one week through several exchanges of letters and a phone call. Finally Pauli essentially conceded Fierz's point in a long letter of February 12, 1957, that also contains first hints at Pauli's psychological resistances (Pauli's emphasis):[59]

> "Your presentation creates in me a feeling of 'formal boredom', to which the fusillade of laughter was of a *compensatory* nature."

This is a curious episode and not easy to understand. Pauli's point seems to have been that he wanted to maintain the particle-antiparticle distinction *independently* of parity, whereas Fierz pointed out that the two-component theory provided no corresponding structural element: In Weyl's form the operations $\mathcal{C}$ and $\mathcal{P}$ simply do not exist separately, in the Majorana form $\mathcal{P}$ exists

---

[57] German original (Pauli, 1990, p. 234, note 54, full sentence): "Indessen sind diese Wellengleichungen, wie ja aus ihrer Herleitung hervorgeht, nicht invariant gegenüber Spiegelungen (Vertauschung von links und rechts) und infolge dessen sind sie auf die physikalische Wirklichkeit nicht anwendbar." The conclusion concerning the non-applicability to physical reality is cancelled in the 1958 edition (cf. Pauli, 1990, p. 150).

[58] German original (Meyenn, 2005a, p. 179): "Lieber Herr Fierz! Ihr Brief vom 6. ist der größte Bock den Sie im Laufe Ihres Lebens geschossen haben! (Wahrscheinlich kommt heute Nachmittag schon eine Berichtigung von Ihnen.) Habe nur den ersten Absatz Ihres der Anstalt entsprungenen Briefes gelesen und mich geschüttelt vor lachen. ... Wenn dieser Brief ankommt (Ihren rahme ich ein!), wissen Sie wohl schon alles!"

[59] German original (Meyenn, 2005a, p. 197): "Ihre Darstellung erzeugt bei mir das Gefühl der 'formalistischen Langeweile', zu der die Lachsalve *kompensatorisch* war."

and $\mathcal{C}$ is the identity (hence not distinguishing). Psychologically speaking, Pauli's point becomes perhaps more understandable if one takes into account the fact that since the fall of 1956 he was thinking about the question of lepton-charge conservation. Intuitively he had therefore taken as self-evident that opposite helicities also corresponded to the particle-antiparticle duality (cf. Meyenn, 2005a, p. 179), even though this association did not correspond to anything in the equations. In a letter dated February 15, 1957, he offered the following in-depth psychological explanation to Fierz (Pauli's emphases):[60]

> "Well, the fusillade of laughter occurred with the term 'Majorana theory' of your first letter. After this catchword I could not go on reading. The immediate association with Majorana clearly has been this: 'aha, particles and antiparticles should no longer exist, one intends to take them away from me (as one takes away a symbol from somebody)!' This causes me *anxiety.* I also know that since last fall the conservation of lepton charge in physics was tremendously important to me – looked upon rationally probably *too important.* I am *anxious* it could turn out to be incorrect and, psychologically speaking, 'discontent' is a euphemism for anxiety. The CP ($\equiv$ Majorana P + exchange between electron and positron) invariance is also important to me, but less so than the conservation of lepton charge. It is *certainly true* that it 'hit upon my Platonic mirror complex'. Particles and antiparticles are the symbol for that more general mirroring (I am not sure to what extent it is particularly *Platonic*). … Mirroring is also a *gnostic symbol* for life and death. There light is *extinguished* at birth and *lightened up* at death. … Obviously, for me the 'mirroring complex' has something to do with death and immortality. *Hence* the anxiety! If the relation between the sleeping mirror image and the one awake would be disturbed, or if they would even be identical (Majorana), then, psychologically speaking, there would neither be life (birth) nor death."

---

[60] German original (Meyenn, 2005a, p. 225): "Also die 'Lachsalve' erfolgte beim Wort 'Majorana Theorie' Ihres ersten Briefes, ich konnte nach diesem Stichwort nicht mehr weiterlesen. Die unmittelbare Assoziation zu Majorana war natürlich 'aha, Teilchen und Antiteilchen soll es nicht mehr geben, die will man mir wegnehmen (wie man jemandem ein Symbol wegnimmt)!' Davor habe ich *Angst.* Ich weiss auch, daß mir schon seit Herbst die Erhaltung der Leptonladung in der Physik ungeheuer wichtig ist – rational betrachtet, vielleicht *zu wichtig.* Ich habe *Angst*, sie könnte sich als unrichtig herausstellen und, psychologisch gesehen, ist 'Unzufriedenheit' ein Euphemismus für Angst. Die CP - ($\equiv$ Majorana P + Vertauschung von Elektron und Positron) Invarianz ist mir auch wichtig, aber weniger wichtig als die Erhaltung der Leptonladung. Es ist *sicher wahr*, daß 'mein platonischer Spiegelkomplex angestochen' war. Teilchen und Antiteilchen sind das Symbol für jene allgemeine Spiegelung (wie weit sie speziell *platonisch* ist, dessen bin ich nicht sicher). … Offenbar hat der 'Spiegelungskomplex' bei mir etwas mit Tod und Unsterblichkeit zu tun. *Daher* die Angst! Wäre die Beziehung zwischen dem schlafenden Spiegelbild und dem Wachenden gestört, oder wären sie gar identisch (Majorana), so gäbe es, psychologisch gesprochen, weder Leben (Geburt) noch Tod."

Fierz later commented on that episode in a personal letter to Norbert Straumann, parts of which are quoted in Straumann (1992).

3.8 CPT and β-Decay, Pauli Group, Cosmological Speculations

3.8.1 CPT: Invariance of Charge, Parity, and Time

In 1955 a collection of essays (Pauli, 1955) by distinguished physicists appeared to celebrate Niels Bohr's 70th birthday. Pauli's contribution, entitled "Exclusion Principle, Lorentz Group and Reflection of Space–Time and Charge", contains the following remarks (Pauli, 1955, pp. 30f):

> "After a brief period of spiritual and human confusion, caused by provisional restriction to 'Anschaulichkeit', a general agreement was reached following the substitution of abstract mathematical symbols, as for instance psi, for concrete pictures. Especially the concrete picture of rotation has been replaced by mathematical characteristics of the representations of the group of rotations in three dimensional space. This group was soon amplified to the Lorentz group in the work of Dirac. ... The mathematical group was further amplified by including the reflections of space and time. ... I believe that this paper also illustrates the fact that a rigorous mathematical formalism and epistemological analysis are both indispensable in physics in a complementary way in the sense of Niels Bohr. While I try to use the former to connect all mentioned features of the theory with help of a richer 'fullness' of plus and minus signs in an increasing 'clarity', the latter makes me aware that the final 'truth' on the subject is still 'dwelling in the abyss'."[61]

This paper of Pauli's can be seen as a follow-up to his spin-statistics paper discussed above, the main difference being that Pauli now considers *interacting* fields. Pauli now *assumes* (1) the validity of the spin-statistics correlation for interacting fields (for which there was no proof at the time), (2) invariance under (the universal cover of) the proper orthochronous Lorentz group SL(2,$\mathbb{C}$) (as in the spin-statistics paper), and (3) locality of the interactions (i.e. involving only finitely many derivatives). Then Pauli shows that this suffices to derive the so-called CPT theorem that states that the combination of charge conjugation (C) and spacetime reflection (PT) is a symmetry.[62]

At the time when Pauli wrote his paper (1955) it was not known whether any of the operations of C, P, or T would separately *not* be a symmetry. This

---

[61] Here Pauli sets the following footnote: "I refer here to Bohr's favorite verses of Schiller: 'Nur die Fülle führt zur Klarheit / Und im Abgrund wohnt die Wahrheit'."

[62] Pauli used a now outdated terminology: instead of CPT he used SR (strong reflection), instead of PT he used WR (weak reflection), and instead of C he used AC (antiparticle conjugation). Preliminary versions of the CPT theorem appeared in papers by Julian Schwinger (1951) and Gerhard Lüders (1954) to which Pauli refers. Two years after Pauli's 1955 paper Res Jost (1957) gave a very elegant proof in the framework of axiomatic quantum field theory.

changed when the experiments of Wu *et al.* in January 1957 showed explicit violations of P and C in processes of β-decay, following a suggestion that this should be checked by Lee and Yang (1956). Pauli had still offered a bet that this would not happen on January 17, 1957 (Pauli's emphases):[63]

> "I do *not* believe *that God is a weak left-hander* and would be prepared to bet a high amount that the experiment will show a symmetric angular distribution of the electrons (mirror symmetry). For I cannot see a logical connection between the *strength* of an interaction and its mirror symmetry."

In view of this firm belief in symmetry the following is remarkable: In his CPT paper Pauli takes great care to write down the most general ultralocal (i.e. no derivatives) four-fermion interaction (for the neutron, proton, electron and neutrino), which is *not* P invariant. In contains 10 essentially different terms with ten coupling constants $C_1, \cdots C_{10}$, only the first five of which are parity invariant (scalars), whereas the other five are pseudoscalars, i.e change sign under spatial inversions. Apparently he did this for the sake of mathematical generality without any physical motivation, as he explicitly stated in a letter to Wu dated January 19, 1957 (Meyenn, 2005a, p. 89; Pauli's emphases):

> "When I considered such formal possibilities in my paper in the Bohr-Festival Volume (1955), I did not think that this could have something to do with nature. I considered it merely as a mathematical play, and, as a matter of fact, I did not believe in it when I read the paper by Yang and Lee ... What prevented me *until now* from accepting this formal possibility is the question of why this restriction of mirroring appears only in the 'weak' interactions, not in the strong ones. *Theoretically*, I do not see any interpretation of this fact, which is empirically so well established."

Lee and Yang took this possibility more seriously: In an appendix to their paper they also wrote down all ten terms for the full, parity non-invariant interaction (Lee and Yang, 1956, p. 258), without any citation of Pauli.

Pauli first learned that the experiments by Wu *et al.* had led to an asymmetric angular distribution from a letter by John Blatt from Princeton, dated January 15, 1957. There Blatt wrote (Meyenn, 2005a, p. 74):

> "I don't know whether anyone has written you as yet about the sudden death of parity. Miss Wu has done an experiment with β-decay of oriented Co nuclei which shows that parity is *not* conserved in β decay. ... We are all rather shaken by by the death of our well-beloved friend, parity."

Pauli, too, was shocked as he stated in his famous letter to Weisskopf dated January 27/28, 1957 (Meyenn, 2005a, pp. 121–127). In that very same letter Pauli already started speculating how symmetry could be restored by letting

---

[63] German original (Meyenn, 2005a,p. 82): "Ich glaube aber *nicht, daß der Herrgott ein schwacher Linkshänder* ist und wäre bereit hoch zu wetten, daß das Experiment symmetrische Winkelverteilung der Elektronen (Spiegelinvarianz) ergeben wird. Denn ich sehe keine logische Verbindung von *Stärke* einer Wechselwirkung und ihrer Spiegelinvarianz."

the constants C_i become dynamical fields, scalar fields for $i = 1, \cdots, 5$ and pseudo-scalar ones for $i = 6, \cdots, 10$: [64]

> "Let us imagine, for example, the terms with $C_1, \cdots, C_5$ being multiplied with a scalar field $\phi(x)$, the terms $C_6, \cdots, C_{10}$ multiplied with a pseudo-scalar field $\hat{\phi}(x)$. For God himself, who can change the sign of $\hat{\phi}(x)$, such a theory would be left-right-invariant – not for us mortal men, however, who do not know anything about that new hypothetical field, except that it is practically constant in space and time on earth (static-homogeneous), and who do not yet[65] have any means to change it."

The mechanism envisaged here to restore symmetry is just that discussed in Section 2.2, where non-dynamical background structures, Σ, are (formally) turned into dynamical quantities, Φ.

3.8.2 The Pauli Group

Since fall of 1956 Pauli's thinking about β-decay was dominated by the lepton-charge conservation. In a paper submitted on March 14, 1957, entitled "On the Conservation of Lepton Charge" Pauli (1957b) once more showed his mastery of symmetry considerations while keeping everything at the largest possible degree of generality.

He starts by considering the most general ultralocal four-fermion interactions (not necessarily preserving parity or lepton charge) in which the neutrino field is represented by a Dirac four-spinor, ψ. For what follows it is convenient to think of the four components of ψ as comprising the following four particle states (per momentum): a left-handed neutrino, ψ_L, a right-handed neutrino, ψ_R, and their antiparticles ψ_L^c and ψ_R^c respectively. Note that this means $\psi_{L,R}^c := (\psi_{L,R})^c$ and that accordingly ψ_L^c is right- and ψ_R^c is left-handed. Here we follow the convention of Kemmer *et al.* (1959).

Next Pauli considers a four-parameter group of canonical transformations (i.e. they leave the anticommutation relations between the fermion fields invariant) of the neutrino field, henceforth called the *Pauli group*, whose interpretation will be given below. These transformations define a symmetry of the interaction-free equations of motion (assuming a massless neutrino throughout), but will generally *not* define a symmetry once the interaction is taken into account. Rather, the following is true (Nishijima, 2004): Suppose that the general interaction depends on a finite number of coupling constants c_i for

---

[64] German original (Meyenn, 2005a, pp. 122–123): "Denken wir uns z.B. die Terme mit $C_1, \cdots, C_5$ mit einem Skalarfeld $\phi(x)$, die Terme mit $C_6, \cdots, C_{10}$ mit einem Pseudo-Skalarfeld $\hat{\phi}(x)$ multipliziert. Für den Herrgott, der das Vorzeichen von $\hat{\phi}(x)$ umdrehen kann, wäre eine solche Theorie natürlich rechts-links-invariant – nicht aber für uns sterbliche Menschen, die wir gar nichts wissen über jenes hypothetische neue Feld, außer daß es praktisch auf der Erde raum-zeitlich konstant (statisch-homogen) ist, und die wir noch kein Mittel haben, es zu ändern."

[65] The "yet" is incorrectly omitted in the official translation (Meyenn, 2005a, p. 126).

$i = 1, \cdots, n$ and that the equations of motion follow from an action principle with Lagrange density $\mathbf{L}\{\,\Sigma \mid \Phi\,\}$, where Σ represents the array of coupling constants (we notationally ignore other non-dynamical structures here) and Φ the dynamical fields. Then the Pauli group acts as covariance in a slightly stronger sense than (5), namely so that

$$\mathbf{L}\{\, g\cdot\Sigma \mid g\cdot\Phi \,\} = \mathbf{L}\{\,\Sigma \mid \Phi\,\}\,. \tag{42}$$

This means that on the level of the Lagrange density (or the Hamiltonian), and hence in particular at the level of the equations of motion, the transformation of the dynamical fields can be compensated for by a transformation of the coupling constants. A large part of Pauli's paper is actually devoted to the determination of that compensating action of the Pauli group on the array of coupling constants.

Next suppose the initial state is chosen to be invariant under the Pauli group, i.e. $g\cdot\Phi = \Phi$ for all g. Then (42) implies that its evolution with interaction parametrized by Σ (the array of c_i's) is identical to the evolution parametrized by $g\cdot\Sigma$ for any g. Hence the outcome of the evolution can only depend on the c_i's through their Pauli-invariant combinations.[66] In particular, since the neutrinoless double β-decay simply has no initial neutrino, this reasoning can be applied to it. If this lepton-charge conservation violating process is deemed impossible, the corresponding Pauli-invariant combination of coupling constants to which the scattering probability is proportional[67] must vanish. This, in turn, gives the sought-after constraint on the possible four-fermion interaction. For (massless) neutrinos in Majorana representation Pauli finally arrived at the result that either only the left- or the right-handed component enters the interaction. This clever sort of reasoning was shortly before used by Pursey (1957) in a less general setting in which the interaction was specialized *a priori* to conserve lepton charge.[68]

More on the history of the search for the right form of the four-fermion interaction may be found in a paper by Straumann (1992). It should also be

---

[66] For illustrative purposes we argue here as if all fields were classical and obeyed classical equations of motion, though Pauli clearly considered quantum theory where the fields become operators. The principal argument is the same, though it makes a big difference between the classical and the quantum case that in the latter we can more easily ascertain the existence of invariant initial states. This is because in quantum theory, assuming there are no superselection rules at work, the superposition principle always allows us to construct invariant initial states by group-averaging any given state over the group (which is here compact, so that the averaging is unambiguously defined). Such states would, for example, appropriately represent physical situations where those observables that distinguish between the states in the group orbit are not measured, may it be for reasons of practice or of principle.

[67] It will be a quadratic combination in leading order of perturbation theory. Explicit calculations had been done by Pauli's assistant Charles Enz (1957).

[68] In terms of the Pauli group, Pursey did not consider the $U(1)$ part.

mentioned that the possibility of neutrinoless double β-decays is currently still under active experimental investigation at the National Gran Sasso Laboratory, where the 2003–2005 CUORICINO experiment set upper bounds for the Majorana mass of the electron neutrino well below one electron Volt (eV). The upcoming next-generation experiment, CUORE, is designed to lower this bound to $0.05\,eV$; compare Gorla (2008).

What is the interpretation of the Pauli group ? Mathematically it is isomorphic to $U(2)$, the group of 2×2 unitary matrices acting on a two-dimensional complex vector space. Here there are two such spaces (per 4-momentum) in which it acts: the "left-handed subspace" that is spanned by the two left-handed components ψ_L and ψ^c_R, and the "right-handed subspace" that is spanned by the two right-handed components ψ_R and ψ^c_L. The two actions of $U(2)$ in these spaces are complex conjugate to each other (see equation (43)). Usually one thinks of the Pauli group as $U(1)\times SU(2)$, which is a double cover of $U(2)$, so that the four real parameters are written as a phase $\exp(i\alpha)$, parametrizing $U(1)$, and two complex parameters a, b satisfying $|a|^2+|sb|^2=1$, which give three real parameters when split into real and imaginary part and which parametrize a three-sphere that underlies $SU(2)$ as group manifold. In this parametrization the action of the Pauli group reads (an asterisk stands for complex conjugation):[69]

$$\begin{pmatrix}\psi_L\\ \psi^c_R\end{pmatrix} \mapsto \exp\bigl(+i\alpha\bigr)\begin{pmatrix}a & b\\ -b^* & a^*\end{pmatrix}\begin{pmatrix}\psi_L\\ \psi^c_R\end{pmatrix}\,, \tag{43a}$$

$$\begin{pmatrix}\psi^c_L\\ \psi_R\end{pmatrix} \mapsto \exp\bigl(-i\alpha\bigr)\begin{pmatrix}a^* & b^*\\ -b & a\end{pmatrix}\begin{pmatrix}\psi^c_L\\ \psi_R\end{pmatrix}\,. \tag{43b}$$

Invariance under the Pauli group is now seen to correspond to an ambiguity in the particle-antiparticle distinction. This ambiguity would only be lifted by interactions that allowed to distinguish the two left and the two right states respectively. In the absence of such interactions the various definitions of "particle" and "antiparticle" are physically indistinguishable, so that the Pauli group acts by *gauge* symmetries in the sense of Section 2.3.

Also, the different presentations of the two-component theory, already discussed in Section 3.7 can be seen here. The Majorana condition reads $\psi=\psi^c$, which in terms of the four components introduced above leads to $\psi_L=\psi^c_R$ and $\psi_R=\psi^c_L$. This can be read in two different ways, depending on whether one addresses ψ_L, ψ^c_L or ψ_L, ψ_R as independent basic states. In the first case one would say that there is a left-handed neutrino and its right-handed antiparticle, whereas in the second case one regards the tuple (ψ_L, ψ_R) as, respectively,

---

[69] Usually the Pauli group is written in terms of the four-component neutrino field ψ as $\psi \mapsto \exp(i\alpha\gamma_5)(a\psi+b\gamma_5\psi^c)$, where $\psi^c := i\gamma_2\psi^*$ is the charge conjugate field. But this is easily seen to be equivalent to (43) if one sets $\psi_{L,R}=\frac{1}{2}(1\pm\gamma_5)\psi$ and $\psi^c_{R,L}=\frac{1}{2}(1\pm\gamma_5)\psi^c$. The more explicit form (43) is better suited for the interpretational discussion (cf. Kemmer *et al.*, 1959). The two-to-one homomorphism from $U(1)\times SU(2)$ to $U(2)$ is given by $\bigl(\exp(i\alpha)\,,\,A\bigr)\mapsto \exp(i\alpha)A$ whose kernel is $\bigl\{(1,\mathbf{1})\,,\,(-1,-\mathbf{1})\bigr\}$.

the left- and right-handed components of a single particle which is identical to its antiparticle.

Beyond weak interaction and β-decay, the Pauli group played a very important role in Pauli's brief participation in Heisenberg's program for a unified field theory. It was Pauli who first showed that the (so far classical) non-linear spinor equation proposed by Heisenberg was invariant under the Pauli group (cf. Heisenberg's account in his letter to Zimmermann from January 7, 1958, in Meyenn, 2005b, p. 779). In this new context the $U(1)$ part of the Pauli group was connected to the conservation of baryon charge and the $SU(2)$ part acquired the meaning of isospin symmetry.[70] The central importance of isospin for this program may already be inferred from the title of the proposed common publication by Heisenberg and Pauli: "On the Isospin Group in the Theory of Elementary Particles". However, due to Pauli's later retreat from this program, the manuscript (cf. Meyenn, 2005b, pp. 849–861) for this publication never grew beyond the stage of a preprint.

3.8.3 Cosmological Speculations

In his last paper on the subject of discrete symmetries, entitled "The Violation of Mirror-Symmetries in the Laws of Atomic Physics",[71] Pauli comes back to the question which bothered him most: How is the strength of an interaction related to its symmetry properties ? He says that having established a violation of C and P symmetry for weak interactions, we may ask why they are maintained for strong and electromagnetic interactions, and whether the reason for this is to be found in particular properties of these interactions. He ends with some speculations on possible connections between violations of C and P symmetry in the laws of microphysics on one hand, and properties of theories of gravitation and its cosmological solutions on the other:[72]

> "Second, one can try to find and justify a connection between symmetry violation in the small with properties of the universe at large. But this exceeds the capabilities of the presently known theory of gravity. ... New ideas are missing to go beyond vague speculations. But this shall not be taken as a definite expression of the impossibility of such a connection."

---

[70] The non-linear spinor equation was at that stage not designed to include weak interaction.

[71] German original (Pauli, 1958): "Die Verletzung von Spiegelungs-Symmetrien in den Gesetzen der Atomphysik".

[72] German original (Pauli, 1958a, p. 4): "Zweitens kann man versuchen, einen Zusammenhang der Symmetrieverletzungen in Kleinen mit Eigenschaften des Universums im Grossen aufzufinden und zu begründen. Dies überschreitet aber die Möglichkeiten der jetzt bekannten Theorien der Gravitation. ... Um bei der Frage des Zusammenhangs zwischen dem Kleinen und dem Grossen über vage Spekulationen hinauszugelangen, fehlen daher noch wesentlich neue Ideen. Hiermit soll jedoch nicht die Unmöglichkeit eines solchen Zusammenhanges bestimmt behauptet werden."

It may be of interest to contrast this expression of a certain open-mindedness for speculations concerning the physics of elementary particles on one side and large-scale cosmology on the other with a more critical attitude in Pauli's very early writings. In Section 65 of his relativity article, where Pauli discussed Weyl's attempt for a unifying theory of gravity and electromagnetism (to which Pauli himself actively contributed), he observed that in Weyl's theory (as well as in Einstein's own attempts from that time) it is natural to suspect a relation between the size of the electron and the size (mean curvature radius) of the universe. But then he comments somewhat dismissively that this "might seem somewhat fantastic" (Pauli, 1958a, p. 202).[73]

4 Conclusion

I have tried to display some of the aspects of the notion of symmetry in the work of Wolfgang Pauli which to me seem sufficiently interesting in their own right. In doing this I have drawn freely from Pauli's scientific œvre, irrespectively of whether the particular part is commonly regarded as established part of present-day scientific knowledge or not. Pauli's faith in the explanatory power of symmetry principles clearly shows up in all corners of his œvre, but it also appears clearly rooted beyond the limits of his science.

In the editorial epilogue to the monumental collection of Pauli's scientific correspondence, Karl von Meyenn reports that many physicists he talked to at the outset of his project spoke against the publication of those letters that contained ideas which did not stand the test of time (Meyenn, 2005b, p. 1375). Leaving aside that this must clearly sound outrageous to the historian, it is, in my opinion, also totally misguided as far as the scientific endeavor is concerned. Science is not only driven by the urge to *know* but also, and perhaps most importantly, by the urge to *understand.* No one who has ever actively participated in science can deny that.

One central aspect of scientific understanding, next to offering as many as possible alternative and complementary explanations for the actual occurrences in nature, is to comprehend why things could not be different from what they appear to be. The insight into a theoretical or an explanatory failure can be as fruitful as an experimental failure. What makes Pauli a great scientist, amongst other reasons, is not that he did not err – such mortals clearly do not exist – , but that we can still learn much from where he erred and how he erred. In that sense, let me end by the following words from Johann Wolfgang von Goethe's "Maximen und Reflexionen" (# 1292):

> "Wenn weise Männer nicht irrten, müßten die Narren verzweifeln."
> ("If wise men did not err, fools should despair.")

---

[73] German original (Pauli, 200b), p. 249): "...was immerhin etwas phantastisch erscheinen mag."

Acknowledgments

I sincerely thank Harald Atmanspacher, Karl von Meyenn, Ulrich Müller-Herold, Reinhard Nesper, and Hans Primas for the invitation to talk at the conference on *Wolfgang Pauli's Philosophical Ideas and Contemporary Science* on the Monte Verità in Ascona, Switzerland.
I also wish to express my strongest appreciation to the editor of Pauli's Scientific Correspondence, Karl von Meyenn, without whose admirable editorial work we would not be in the position to share many of Wolfgang Pauli's wonderful insights. Thank you!

References

Alexandrov, A.D. (1975): Mappings of spaces with families of cones and space-time transformations. *Annali di Matematica Pura ed Applicata* **103**, 229–257.

Anderson, J.L. (1967): *Principles of Relativity Physics.* Academic Press, New York.

Arens, R. (1971): Classical Lorentz invariant particles. *Journal of Mathematical Physics* **12**, 2415–2422.

Bacry, H. (1967): Space-time and degrees of freedom of the elementary particle. *Communications in Mathematical Physics* **5**, 97–105.

Bellac, M. le and Lévy-Leblond, J.M. (1973): Galilean electromagnetism. *Nuovo Cimento* **14B**, 217–233.

Bohr, N., Kramers, H.A. and Slater, J.C. (1924): Über die Quantentheorie der Strahlung. *Zeitschrift für Physik* **24**, 69–87.

Bopp, F. and Haag, R. (1950): Über die Möglichkeit von Spinmodellen. *Zeitschrift für Naturforschung* **5a**, 644–653.

Brown, H. (2005): *Physical Relativity: Space-Time Structure from a Dynamical Perspective.* Oxford University Press, Oxford.

Cantor, G. (1895): Beiträge zur Begründung der transfiniten Mengenlehre. (Erster Artikel). *Mathematische Annalen* **46**, 481–512.

Case, K.M. (1957): Reformulation of the Majorana theory of the neutrino. *Physical Review* **107**, 307–316.

Coleman, R.A. and Korte, H. (1980): Jet bundles and path structures. *Journal of Mathematical Physics* **21**, 1340–1351.

Coleman, S. and Mandula, J. (1967): All possible symmetries of the S matrix. *Physical Review* **159**, 1251–1256.

Ehlers, J. and Köhler, E. (1977): Path structures on manifolds. *Journal of Mathematical Physics* **18**, 2014–2018.

Einstein, A. (1914a): Prinzipielles zur verallgemeinerten Relativitätstheorie und Gravitationstheorie. *Physikalische Zeitschrift* **15**, 176–180.

Einstein, A. (1914b): Die formale Grundlage der allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlichen Preussischen Akademie der Wissenschaften* **1914**, 1030–1085.

Einstein, A. (1918): Prinzipielles zur allgemeinen Relativitätstheorie. *Annalen der Physik* **55**, 241–244.

Einstein, A. (1920): Diskussionsbemerkung. *Physikalische Zeitschrift* **21**, 662.

Einstein, A. (1925): Elektron und allgemeine Relativitätstheorie. *Physica* **5**, 330–334.

Einstein, A. and Grossmann, M. (1914a): Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation. *Zeitschrift für Mathematik und Physik* **62**, 225–261.

Einstein, A. and Grossmann, M. (1914b): Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie. *Zeitschrift für Mathematik und Physik* **63**, 215–225.

Einstein, A. and Pauli, W. (1943): On the non-existence of regular stationary solutions of relativistic field equations. *Annals of Mathematics* **44**, 131–137.

Enz, C.P. (1957): Fermi interaction with non-conservation of 'Lepton charge' and of parity. *Nuovo Cimento* **6**, 250–254.

Enz, C.P. (2002): *No Time to Be Brief.* Oxford University Press, Oxford.

Enz, C.P. and Meyenn, K. von (eds.) (1994): *Wolfgang Pauli. Writings on Physics and Philosophy.* Springer, Berlin.

Fierz, M. (1939): Über die relativistische Theorie kräftefreier Teilchen mit beliebigem Spin. *Helvetica Physica Acta* **12**, 3–37.

Fierz, M. and Pauli, W. (1939): On relativistic wave equations for particles of arbitrary spin in an electromagnetic field. *Proceedings of the Royal Society (London)* **A173**, 211–232.

Fierz, M. and Weisskopf, V.F. (eds.) (1960): *Theoretical Physics in the Twentieth Century. A Memorial Volume to Wolfgang Pauli.* Interscience, New York.

Fushchich, W. and Nikitin, A. (1979): On the new invariance group of Maxwell equations. *Lettere al Nuovo Cimento* **24**, 220–224.

Fushchich, W. and Shtelen, V. (1991): Are Maxwell's equations invariant under the Galilei transformations ? (in Russian). *Doklady Akademii Nauk Ukrainy* **A3**, 22–26.

Fushchich, W., Shtelen, V. and Serov, N.I. (1993): *Symmetry Analysis and Exact Solutions of Equations of Nonlinear Mathematical Physics.* Kluwer, Dordrecht.

Giulini, D. (1995): Asymptotic symmetry groups of long-ranged gauge configurations. *Modern Physics Letters* **A10**, 2059–2070.

Giulini, D. (2006): Algebraic and geometric structures in special relativity. In: Lämmerzahl, C., and Ehlers, J. (eds.), *Special Relativity: Will it Survive the Next 101 Years ?* Springer, Berlin, pp. 45–111.

Giulini, D. (2007a): Some remarks on the notions of general covariance and background independence. In: Seiler, E., and Stamatescu, I.O. (eds.), *Approaches to Fundamental Physics.* Springer, Berlin, pp. 105–120.

Giulini, D. (2007b): Electron spin or "classically non-describable two-valuedness". `http://arxiv.org/pdf/0710.3128`.

Gödel, K. (1949): An example of a new type of cosmological solutions of Einstein's field equations of gravitation. *Reviews of Modern Physics* **21**, 447–450.

Goldin, G.A. and Shtelen, V.M (2001): On Galilei invariance and nonlinearity in electrodynamics and quantum mechanics. *Physics Letters* **A279**, 321–326.

Goldstein, N.J. (2007): Inertiality implies the Lorentz group. *Mathematical Physics Electronic Journal* **13**, paper 2, `http://www.ma.utexas.edu/mpej/MPEJ.html`

Golfand, Yu.A. (1971): Extension of the algebra of Poincaré group generators and violation of P invariance. *Journal of Experimental and Theoretical Physics Letters* **13**, 323–326.

Gorla, P. (2008): The CUORE experiment: Bolometric detectors in 1-ton scale projects. *Journal of Low Temperature Physics* **151**, M04.

Gross, D. and Perry, M. (1983): Magnetic monopoles in Kaluza-Klein theories. *Nuclear Physics* **B115**, 29–48.

Hawking, S.W., King, A.R. and McCarthy, P.J. (1976): A new topology for curved space-time which incorporates the causal, differential, and conformal structure. *Journal of Mathematical Physics* **17**, 174–181.

Hermann, A., Meyenn, K. von and Weisskopf, V.F. (eds.) (1979): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel. Band I: 1919–1929.* Springer, New York.

Joos, E., Zeh, H.D., Kiefer, C., Giulini, D., Kupsch, J. and Stamatescu, I.O. (2003): *Decoherence and the Appearence of a Classical World in Quantum Theory.* Springer, Berlin.

Jordan, P. and Pauli, W. (1928): Zur Quantenelektrodynamik ladungsfreier Felder. *Zeitschrift für Physik* **47**, 151–173.

Jost, R. (1957): Eine Bemerkung zum CTP Theorem. *Helvetica Physica Acta* **30**, 409–416.

Kemmer, N., Polkinghorne, J. and Pursey, D. (1959): Invariance in elementary particle physics. *Reports on Progress in Physics* **22**, 368–432.

Klein, F. (1872): *Vergleichende Betrachtungen über neuere geometrische Forschungen.* Verlag von Andreas Deichert, Erlangen. Reprinted in: *Mathematische Annalen* **43**, 63-100, 1892.

Kretschmann, E. (1917): Über den physikalischen Sinn der Relativitätspostulate, A. Einsteins neue und seine ursprüngliche Relativitätstheorie. *Annalen der Physik* **53**, 575–614.

Kronig, R. and Weisskopf, V.F. (eds.) (1964): *Collected Scientific Papers by Wolfgang Pauli.* Interscience, New York.

Lüders, G. and Zumino, B. (1958): Connection between spin and statistics. *Physical Review* **110**, 1450–1453.

Lee, T.D. and Yang, C.N. (1956): Question of parity conservation in weak interactions. *Physical Review* **104**, 254–258.

Lorentz, H.A. (1914): *Das Relativitätsprinzip. Drei Vorlesungen, bearbeitet von W.H. Keesom.* Teubner, Leipzig.

Mach, E. (1933): *Die Mechanik in ihrer Entwicklung.* 9. Auflage. Brockhaus, Leipzig.

McLennan, J.A. (1957): Parity nonconservation and the theory of the neutrino. *Physical Review* **106**, 821–822.

Meyenn, K. von (1987): Pauli's belief in exact symmetries. In: Doncel, M.G., Hermann, A., Michel, L., and Pais, A. (eds.), *Symmetries in Physics (1600–1980).* Servei de Publicaciones, Bellaterra (Barcelona), pp. 329–358.

Meyenn, K. von (ed.) (1985): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band II: 1930–1939.* Springer, Berlin.

Meyenn, K. von (ed.) (1993): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band III: 1940–1949.* Springer, Berlin.

Meyenn, K. von (ed.) (1996): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil I: 1950–1952.* Springer, Berlin.

Meyenn, K. von (ed.) (1999): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil II: 1953–1954.* Springer, Berlin.

Meyenn, K. von (ed.) (2001): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil III: 1955–1956.* Springer, Berlin.

Meyenn, K. von (ed.) (2005a): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil IV-A: 1957*. Springer, Berlin.

Meyenn, K. von (ed.) (2005b): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil IV-B: 1958*. Springer, Berlin.

Minkowski, H. (1909): *Raum und Zeit*. Teubner, Leipzig.

Nishijima, K. (2004): Pauli and the Pauli group. *Nuclear Physics B – Proceedings Supplements* **137**, 1–4.

O'Raifeartaigh, L. (1965): Lorentz invariance and internal symmetry. *Physical Review* **139**, B1052–B1062.

O'Raifeartaigh, L. and Straumann, N. (2000): Gauge theory: Historical origins and some modern developments. *Reviews of Modern Physics* **72**, 1–23.

Pauli, W. (1919a): Merkurperihelbewegung und Strahlenablenkung in Weyl's Gravitationstheorie. *Verhandlungen der Deutschen Physikalischen Gesellschaft* **21**, 742–750.

Pauli, W. (1919b): Zur Theorie der Gravitation und der Elektrizität von Hermann Weyl. *Physikalische Zeitschrift* **20**, 457–467.

Pauli, W. (1925): Über den Einfluss der Geschwindigkeitsabhängigkeit der Elektronenmasse auf den Zeemaneffekt. *Zeitschrift für Physik* **31**, 373–385.

Pauli, W. (1927): Zur Quantenmechanik des magnetischen Elektrons. *Zeitschrift für Physik* **43**, 601–623.

Pauli, W. (1939): Über ein Kriterium für Ein- oder Zweiwertigkeit der Eigenfunktionen in der Wellenmechanik. *Helvetica Physica Acta* **12**, 147–168.

Pauli, W. (1940): The connection between spin and statistics. *Physical Review* **58**, 716–722.

Pauli, W. (1941): Relativistic field theories of elementary particles. *Reviews of Modern Physics* **13**, 203–232.

Pauli, W. (ed.) (1955): *Niels Bohr and the Development of Physics*. Pergamon Press, London.

Pauli, W. (1956): Schlußwort durch den Präsidenten der Konferenz "Fünfzig Jahre Relativitätstheorie". *Helvetica Physica Acta, Supplementum* **IV**, 261–267.

Pauli, W. (1957a): Zur älteren und neueren Geschichte des Neutrinos. *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich* **102**, 387–388. Extended version in: W. Pauli: *Aufsätze und Vorträge über Physik und Erkenntnistherorie*. Vieweg, Braunschweig 1961. pp. 156–180.

Pauli, W. (1957b): On the conservation of the lepton charge. *Nuovo Cimento* **6**, 204–215.

Pauli, W. (1958a): Die Verletzung von Spiegelungs-Symmetrien in den Gesetzen der Atomphysik. *Experientia* **14**, 1–5.

Pauli, W. (1958b): *Theory of Relativity*. Pergamon Press, Oxford.

Pauli, W. (1965): Continuous groups in quantum mechanics. *Ergebnisse der exakten Naturwissenschaften* **37**, 85–104.

Pauli, W. (1990): *Die allgemeinen Prinzipien der Wellenmechanik*. Springer, Berlin. Newly edited and annotated by Norbert Strauman.

Pauli, W. (2000a): *Pauli Lectures on Physics. Volume 6. Selected Topics in Field Quantization*. Dover, New York.

Pauli, W. (2000b): *Relativitätstheorie*. Springer, Berlin. Reprint of the original "Encyclopädie-Artikel" from 1921 with various additions, including Pauli's own supplementary notes of 1956. Edited and annotated by Domenico Giulini.

Pfister, H. (2004): Newton's first law revisited. *Foundations of Physics Letters* **17**, 49–64.

Pursey, D. (1957): Invariance properties of Fermi interactions. *Nuovo Cimento* **6**, 266–277.

Serpe, J. (1952): Sur la théorie abrégée des particules de spin $\frac{1}{2}$. *Physica* **18**, 295–306.

Serpe, J. (1957): Remarks on parity conservation and the two-component theory of the neutrino. *Nuclear Physics* **4**, 183–187.

Sexl, R.U. and Urbantke, H.K. (2001): *Relativity, Groups, Particles: Special Relativity and Relativistic Symmetry in Field and Particle Physics*. Springer, Wien.

Sorkin, R. (1983): Kaluza-Klein monopole. *Physical Review Letters* **51**, 87–90.

Straumann, N. (1992): Von der Stech-Jensen-Transformation zur universellen V-A-Wechselwirkung. *Archive for History of Exact Sciences* **44**, 365–386.

Straumann, N. (2004): The role of the exclusion principle for atoms to stars: A historical account. Online available at `arxiv.org/pdf/0403/0403199`.

Weinberg, S. (1995): *The Quantum Theory of Fields. Volume I Foundations*. Cambridge University Press, Cambridge.

Weyl, H. (1928): *Gruppentheorie und Quantenmechanik*. Hirzel, Leipzig.

Weyl, H. (1991): *Raum Zeit Materie*. Springer, Berlin. Based on the fifth edition (1923), with additional appendices, edited and annotated by Jürgen Ehlers.

Wigner, E. (1931): *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren*. Vieweg, Braunschweig.

Zeeman, E.C. (1964): Causality implies the Lorentz group. *Journal of Mathematical Physics* **5**, 490–493.

A New Idea of Reality: Pauli on the Unity of Mind and Matter

William Seager

Department of Philosophy, University of Toronto, Canada,
`seager@utsc.utoronto.ca`

1 Introduction

Here are two representative but somewhat cryptic remarks of Wolfgang Pauli on the relation of mind and matter: "physis and psyche are probably two aspects of one and the same abstract fact" and "…a mirror-image principle is a natural way to give an illustrative representation of the psychophysical relationship."[1] Although the general idea of a "dual-aspect" account of the relation between mind and matter is the obvious content of these remarks, two particular questions spring to mind. The first is, why does Pauli say that mind and matter are aspects of an "abstract" fact and the second is, how seriously does Pauli intend the mirror image analogy. A serious use implies some strong and quite radical conclusions about the nature of mind, which may also help us to understand what Pauli intended by reference to abstract facts.

There are any number of problems and issues that arise in the philosophy of mind but two main strands of thought can be distinguished which characterize Pauli's interests. One is a concern with the *operation* of the mind. This can be regarded as an empirical issue addressed by psychology, but it also has a philosophical aspect. Philosophers have an interest in the dynamics of mental states, the appropriate categorization of mental states and, perhaps most especially, the contents of mental states. A quick caricature of the typical approach to this issue can be given with belief-desire theory. On this view the mind is – to a crude first approximation – described in terms of two core states, one whose function is to represent the way the world is, and the other to represent the way the world should become (according to the subject). Mental dynamics is the evolution of the belief-desire system in light of the continually changing environment as the subject pursues his or her desires.

---

[1] Both passages are from a letter from Pauli to Carl Gustav Jung (Meier, 2001, p. 159).

More sophisticated accounts increase the number of distinct mental states and provide for more fine-grained characterizations of their contents but the basic idea is clear enough.

The belief-desire scheme is very abstract and all sorts of activity fall under it – from the behavior of quite simple animals to that of not so simple physicists. While Pauli was no cognitive scientist he did have a very keen interest in this aspect of the problem of mind which he famously approached from the perspective of C.G. Jung's archetypal psychology. He wrote extensively on this subject. Pauli's (1952) study of the genesis of Kepler's scientific ideas provides a lengthy and deeply worked out example of the application of Jung's ideas to the psychological dynamics involved in scientific theorizing (of which more below).

The second strand of thought is more purely philosophical. It is the metaphysical or ontological question of the ultimate nature of mind and its place in nature. On this topic Pauli had definite views but wrote nothing systematic or extensive. Rather, his writing is fragmentary, largely taking the form of suggestive comments occurring sporadically in writing on other subjects. Nevertheless, it is possible to assemble at least a framework of a theory about the relation of mind and matter which Pauli would endorse (though it is impossible to know in any detail how Pauli would have liked the theory to be extended). This task has been most notably undertaken in Atmanspacher and Primas (2006). Instead of replicating their exemplary efforts, I want to trace out a possible argumentative path from Pauli's understanding of quantum physics to the philosophy of mind. To begin, since Pauli's view falls under the general heading of a dual-aspect theory and such theories have a long history, it is worth briefly exploring their philosophical ancestry.

2 Mind and Body from Descartes to Leibniz

The father of the modern mind-body problem is of course René Descartes. His infamous dualism was the least revisionary response to the new mechanical world view (which he had done so much to bring into prominence). It was evident to Descartes that mental features were necessarily distinct from the physical for a variety of reasons ranging from a traditional commitment to a radical freedom of will to purely metaphysical arguments that the distinctive phenomenological and epistemological properties of consciousness precluded identifying the mental with the physical. Since it was also evident to Descartes that the world could be given an (almost) complete physical description and seemed to act (almost) always in accord with purely physical law, he attempted to isolate the "mechanical interface" between matter and mind to a single locus – the pineal gland. Thus Descartes was able to let the physical world evolve *almost* in accord with purely physical law while retaining the common-sense belief in the causal interaction of mental and physical states.

It was nonetheless instantly recognized that Descartes's theory was metaphysically highly suspicious. The mysterious appearance of large numbers of obscure miracles within human brains – and only *human* brains – is unattractive at best and, strictly speaking, decisively refutes the idea that the physical world is a complete and closed system. It was further argued that the proposed interaction between two utterly distinct substances was simply metaphysically incoherent.[2]

There was no shortage of proposed upgrades and replacements of Cartesian dualism. Three of particular note are the occasionalism of Malebranche, the dualistic parallelism of Leibniz, at least as popularly presented,[3] and the radical monistic parallelism of Spinoza.

Pauli implicitly and rightly dismisses Malebranche's view. Occasionalism posits a perfect forgery of Cartesian interactionism perpetrated by God who intervenes to wiggle, so to speak, the pineal gland at the appropriate times, in the appropriate direction.[4] With respect to the scientific view of the world, occasionalism has all the vices of Descartes's theory and none of its virtues. It endorses the occasional violation of physical law while denying us our commonsense belief in the efficacy and responsiveness of our mental states.

Leibnizian parallelism permits physical nature to form a closed and complete system of causes and effects. The appearance of interactions between the mental and the physical is the product of divine pre-established harmony which synchronizes the two realms. Leibniz's famous illustration is the way two accurate clocks will chime the hours together despite being causally isolated from each other.[5]

Well aware of this episode of philosophical history, Pauli (1954, p. 290; in Enz and Meyenn, 1994, p. 155) was suspicious of parallelism, saying that

---

[2] Of special note here is Descartes's epistolary exchange with Princess Elisabeth in which Descartes is driven to admit that the relation between mind and matter is entirely inexplicable and must be regarded as a "primitive notion" (see Shapiro, 2007).

[3] Leibniz's view was not a two-world or two substance type parallelism but something more akin to a idealist monism (see e.g. Adams 1994, Chap. 9). The monads are fundamentally mental in nature and exhaust reality (even God is "merely" another monad albeit one with very special properties).

[4] Malebranche's view was of course not restricted to God wiggling the pineal gland – all causal relations had to be mediated by God since only He possessed actual causal power (for a good overview see Schmaltz, 2006).

[5] It is interesting to recall that Leibniz must have been aware of Huygens's (the man who introduced Leibniz to advanced mathematics) 1665 observation that mechanical clocks can achieve a causally mediated synchronization when they form a coupled system. It is also worth noting that the problem of clock synchronization was of first importance throughout this period insofar as it was vital for sea navigation and the determination of longitude. It would be interesting to know how or whether such considerations influenced Leibniz's choice of example.

> "ever since the 17th century [the psycho-physical interconnections] have been something of an embarrassment to the world-picture of 'classical' physics, in that it has been necessary to postulate … a connection of a different, 'parallelistic' kind, in addition to the ordinary causal connection."

Presumably, this necessity arises from the requirement that the physical world be in some sense causally closed and complete. We shall see that, on Pauli's understanding of this requirement, determinism is not entailed by closure and completeness.

For Pauli, parallelism was highly unsatisfactory but in the context of classical physics perhaps unavoidable. Two weaknesses of the parallelistic viewpoint were particularly troubling to Pauli. The first is the strange uniqueness of the mental. No other realm forces us to take the parallelist escape hatch. This cries out for an explanation, which leads to the second dissatisfaction: that the mental–physical parallelism, like all unexplained correlations, suggests the existence of an underlying "common cause" or ground of the correlation. Pauli (1954, p. 290; in Enz and Meyenn, 1994, p. 155) writes:

> "Is it *only* in the association of physical and psychical processes, and not in other situations as well, that a parallelistic relation exists? And does not a relation of parallelism mean that it is justifiable to demand that that which is associated, or 'corresponds' (the corresponding) should also be embraced conceptually in a unity of essence?"

As Pauli notes, the natural response to these problems is some form of ontological monism. This is precisely the route that was taken by post-Cartesian philosophy, although the philosophical development was conditioned by a reflexive acceptance of the basic grounds of dualism, i.e., the seemingly evident separation of mind (or consciousness) from matter (or "mere extension").

Nowadays, the most natural and dominant form of monism is materialism or physicalism. However, only a *radical* materialism will eliminate the problem of dualism. Up to the mid 20th century most of what passed for materialism was really what is better called property dualism. The basic materialist position was to deny a duality of substance but endorse the idea that matter could possess mental properties (see Seager, 2007). While clearly anti-Cartesian, such a view is not wholesale materialism.

Radical materialism was, it seems, never taken seriously by Pauli. Why is that? It may be because Pauli was unaware of the nascent radical materialist identity theories that were promulgated in the mid 1950s.[6] More important, as we shall see, its fundamental commitment to scientific realism made (or would have made) the radical materialism of the identity theory a non-starter for Pauli.

---

[6] There is an interesting question here. The work of Smart (1959) and Place (1956) was likely unknown to Pauli (their famous articles on the identity theory were published after Pauli's death). Mainstream philosophy was still hostile to radical materialism up to the 1950s more or less, regarding it as either provably false (Broad, 1925) or even incoherent (Ryle, 1949).

But until the 20th century, materialism was far from the mainstream. Although the need for a monistic viewpoint was recognized, the standard post-Cartesian monism was one in which the mental realm formed the ontological basis of the world, expressed in a variety of idealist philosophies. The philosophical lineage of idealism begins with Leibniz's response to Descartes's interactionist dualism. Leibniz's ontological bedrock – the monad – was intrinsically mentalistic in nature. Idealism was "perfected" by Kant and his successors both on the continent and in the British Isles.

Idealism fails to meet Pauli's *desideratum* however. It utterly fails to provide a conception of a "unified essence" encompassing both mind and matter but rather explicitly aims to reduce the physical to relations amongst states which are fundamentally mental in nature.[7]

3 Pauli a Spinozist ?

But there is another strand of thought in the history of philosophy which begins with the reaction against Descartes by Spinoza, and which does not lead to either materialism or idealism. Spinoza held that the fundamental reality (which he labeled "God", although it bears few if any of the divine properties traditionally recognized in the Abrahamic religious tradition) was an infinite, all-encompassing substance which was in itself neither mental nor physical. This substance possessed an infinity of attributes, only two of which our minds can comprehend: matter and consciousness. Each of the attributes provides a complete representation of God from, as it were, their own perspective. Thus they can be regarded as mirror images of each other.

Spinoza's metaphysical view is usually called dual-aspect theory (but it should be borne in mind that Spinoza allowed for an infinity of aspects). Obviously, it immediately solves the technical problems with Cartesian dualism: there is no peculiar interaction between the mental and physical realms; there is no breakdown in the lawful evolution of the physical world and the evident correlations between physical and mental events are elegantly explained. I think it is fair to say that Spinoza's theory provides a reasonably robust sense in which mind and matter are joined in a "unity of essence".

The main "splintering" of Spinozism occurred in the latter half of the 19th century with the division of the theory into dual-aspect and so-called neutral monistic forms (see Stubenberg, 2005). There is no canonical formulation of the difference between these metaphysical theories. Both accounts accept that there is a neutral, that is, neither mental nor physical in itself, foundation of

---

[7] Kant's position in the history of idealism is slightly uncomfortable, or perhaps merely transitional. Arguably, his notion of the noumenal realm of things-in-themselves provides room for a "unified essence" conception of monism, but there is little evidence that Kant accepted such a position. Post-Kantian idealists solved this problem by essentially eliminating the noumenal, or assimilating it to the mental.

reality. The core difference lies in the issue of reducibility: the neutral monists can be viewed as advocating the view that both mental and physical reduce (in the appropriate way) to the neutral substrate. Dual-aspect theory denies this, instead regarding mind and matter as co-fundamental ways of apprehending the neutral substrate. A second difference is that neutral monists are inclined to think that mind and matter retain a kind of separateness insofar as only certain configurations of the neutral underlie mental states while other, distinct configurations underlie the physical states. Dual-aspect theories cleave to the view that the neutral can be generally and equally regarded under either aspect.

Pauli's view is clearly of the dual-aspect persuasion (see Atmanspacher and Primas, 2006) but the relationship with Spinoza's views is not one of direct influence. Judging from correspondence, mostly with Fierz, it seems Pauli had a troubled relationship with Spinoza, primarily because of Einstein's frequent appeals to Spinozism as both a kind of spiritual foundation for a life-attitude and as a champion of the classical ideal of the physical world being deterministic and "closed" under natural laws. It was not the Spinozistic spiritualism that bothered Pauli, but both of Einstein's scientific desiderata were strongly opposed by Pauli who was not afraid of indeterminism in nature and rejected Einstein's (1953, p. 6) demand that it be recognized

> "there is such a thing as the real state of a physical system, which exists objectively, independently of any observation or measurement, and can in principle be described by the methods of expression of physics."

Nonetheless, there is a strong affinity in the content of Pauli's and Spinoza's views. But the genesis of Pauli's dual-aspect theory fundamentally stems from his appreciation of certain insights provided by quantum theory rather than any study of the history of philosophy. In fact, I think that Pauli's quantum approach adds a new and very interesting argument for the dual-aspect account of the mind-matter relation which makes it of real philosophical interest.

4 The Unconscious in the Context of Discovery

To see how this works, we must turn briefly to the first of the two strands of thought I characterized above, to the issue of the operation and the contents of the mind, and especially how these affect or constrain the genesis of scientific theories. There has been something of a tradition in the philosophy of science to relegate the creative aspect of theory development to the philosophical ghetto of what was called the "context of discovery" (see Reichenbach, 1938). In contrast to the rationally virtuous "context of justification" where logic and proper experimental protocol hold sway, the context of discovery is a free-for-all of imagination, dreams and rampant illogicality. There are no methodological or logical constraints within the domain of discovery. Reichenbach and Popper regarded this as ultimately harmless however, because, no matter their provenance, all scientific hypotheses have to be *tested* as well as *invented*. In the crucible of the laboratory or the observatory, the scientist's dreams will eventually confront reality.

In what surely must have been a conscious echo of this doctrine Pauli (1956, p. 51; in Enz and Meyenn, 1994, p. 138) characterized the difference between discovery and justification in these terms:

> "Mathematics and natural science are specially distinguished ... by being teachable and verifiable. ... By teachability I mean communicability to others of trains of thought and of results made possible by a progressive tradition ... [requiring an] intellectual effort of quite a different kind from that required for the discovery of something new. In the latter process the creative irrational element finds more essential expression ... In science there is no general rule for passing from the empirical material to new concepts and theories ... ."

However, unlike Reichenbach and Popper, Pauli thought it was illuminating and important to the project of understanding science itself to investigate the process of scientific creativity. In this he somewhat anticipated the current attitude to this issue in which philosophers such as Giere (1988) and Thagard (1992) seek to investigate theory creation. But while they regard scientific creativity as simply another aspect of human cognition which can be studied by normal scientific methods, in particular those of the cognitive sciences, Pauli drew on ideas from Jungian archetypal psychology and its account of the unconscious wellsprings of thought.[8]

There is a good deal that could be said about this aspect of Pauli's thought – especially about the role of particular archetypal figures, but I want to just list three somewhat peculiar features of his views and then focus on a fourth which I think is more important for our purposes. The archetypes provide a creative source of fundamental conceptual building blocks that is ultimately grounded in the unconscious mind. Pauli regards them as (1) entirely unquantifiable (does this mean that they are *not* subject to standard scientific investigation, contra modern cognitive science-oriented research ?), (2) completely mysterious in their origin (particularly and explicitly with regard to their putative emergence via biological natural selection – about which Pauli seemed to have general reservations), and (3) possibly essentially linked to various paranormal or parapsychological phenomena, towards which Pauli seems to have had an unfortunately rather uncritical attitude (see Pauli, 1954, pp. 297ff; in Enz and Meyenn, 1994, p. 161ff). None of these perhaps dubious ideas matter when we think about the path Pauli took towards his dual-aspect solution to the mind-body problem.

More interesting is the paradox, as Pauli calls it, which is involved in any attempt to understand or come to know the unconscious mind and its contents. In a discussion of dream interpretation as a possible source of knowledge about

---

[8] The most sustained discussion of how Jungian psychology can be applied to the problem of theory formulation can be found in Pauli's (1952) study of Kepler.

the unconscious Pauli (1954, p. 286; in Enz and Meyenn, 1994, p. 153) writes that

> "the mere apprehension of the dream has already, so to speak, altered the state of the unconscious, and thereby, in analogy with a measurement in quantum physics, created a new phenomenon."

Pauli goes on to illustrate the paradox with two quotations from Jung about the overall structure of the "psyche": "the psyche is a conscious–unconscious whole" and "the psyche and its contents are the only reality which is given to us *without a medium*" (Pauli, 1954, pp. 287–288; in Enz and Meyenn, 1994, p. 153).[9] Pauli addresses the somewhat superficial "paradox" of Jung, seemingly claiming that the unconscious is given to consciousness as part of the conscious-unconscious whole which is the psyche. I tend to think this is merely carelessness of expression on Jung's part and what he intends to emphasize is that whatever comes to consciousness is given in an unmediated form.

Still, there is an obvious problem with the idea that we could ever become conscious of the unconscious as such, for the act would immediately negate the unconscious aspect which we were hoping to investigate. The unconscious would necessarily remain hidden and to the extent we had any access to it at all would be irrevocably altered by our attempt to gain conscious access to it. This does not mean, and Pauli did not mean to imply, that the unconscious is *unknowable*. What it means is that the unconscious is not knowable via the operation of conscious introspection (or bringing to consciousness); it is knowable only by indirect, more abstract or theoretical, means. On the other hand, the conscious mind is knowable directly, via introspection. In addition, it is arguable that there is no non-introspective or non-experiential way to come to know the core feature of conscious mental states: what is commonly referred to as their phenomenal character. Thus the conscious and unconscious "aspects" of the mind are knowable only via distinct methods or operations even though they form a whole or a unity.

The analogy with quantum mechanics is obvious (as Pauli noted).[10] Relative to the psychic whole, conscious and unconscious mentality are comple-

---

[9] Although entirely tangential to the concerns of this article, Jung's second remark is extremely interesting. Is Jung endorsing the old "way of ideas" in which the only thing we are "directly" aware of are our own states of consciousness ? What does he mean by a "medium" and how does he understand the pregnant term "given" ?

[10] There is something of a disanalogy as well. According to the usual interpretation, the conscious and unconscious form two entirely distinct areas of the mind, preexisting within the mind and in some kind of causal interaction. These features have to be rethought if the analogy with quantum mechanics is to be complete and if the model for a dual-aspect theory is to be fully adequate. However, since at this point the invocation of complementarity serves only as an analogy it is perhaps not very important whether or not it can be extended to a complete analogy – it will do its job if it suggests a direction of thought which leads towards the dual-aspect ontology of mind.

mentary observables. It is the idea of complementarity which, I think, underpins Pauli's endorsement of the dual-aspect view of the mind-matter relation and which also suggests a pathway towards that view that is original and remains of current interest. The route from complementarity to dual-aspect monism is not direct. It runs through one of the basic issues in general philosophy of science: the realism debate.

5 Scientific Realism and Anti-Realism

The core idea of scientific realism is that science provides the truth about the ultimate nature of the world as it is in itself. There is a host of different versions of the doctrine[11] but we can proceed with a superficial division into what I will call strong and weak scientific realism (SSR and WSR, respectively). SSR claims that current mature sciences provide an (approximately) true representation of the nature of reality as it is in itself. Of course, defenders of SSR realize that current science is provisional and in flux, but they regard the "main outlines" of reality as having been established by the science we now possess. While current science is incomplete and is both revisable and will be revised, the basic picture of the structure of the world will not change radically. In effect, SSR holds that science *replaces* metaphysics as the arbiter of the traditional question: What is the ultimate nature of reality ? In addition, SSR boldly maintains that this question has, in broad terms, been answered.

SSR is an extreme position which few if any would actually defend. It can be contrasted with the more modest claims of WSR, namely that science *aims* to provide a true representation of the nature of reality as it is in itself. Unlike SSR, WSR regards the question of how close current science is to fulfilling this aim as rather imponderable. It is entirely possible (perhaps a few foundational issues in physics even suggest) that developments in science will occur that conceivably could lead to a radical and very extensive overthrow of our basic understanding of nature. But WSR agrees with SSR that the old philosophical question about the ultimate nature of reality is and should be answered by science rather than metaphysics.

It is my contention that Pauli's rejection of scientific realism (even WSR) is the key to his adoption of a dual-aspect view of mind and that this rejection is driven by his acceptance of complementarity in quantum mechanics and his wish to follow Bohr in extending complementarity beyond physics narrowly construed. It is no surprise to find that Bohr is Pauli's inspiration here and Bohr decisively rejected scientific realism. Here are two representative quotations from Bohr (1934, p. 18):

> "In our description of nature the purpose is not to disclose the real essence of the phenomena but only to track down, so far as it is possible, relations between the manifold of our experience."

---

[11] For characterizations, attacks and defenses see, variously, Fraassen, 1981; Hacking, 1983; Cartwright, 1983; Harré, 1986.

and, very bluntly, "it is wrong to think that the task of physics is to find out how Nature is" (Petersen, 1963, p. 12). And here is one quotation from Pauli (1950, p. 73; in Enz and Meyenn, 1994, pp. 35–36) which echoes Bohr's sentiments and explicitly links the issue to the crucial question of whether science reveals nature as it is in itself or can even take on that task as a legitimate goal:

> "As a result of the development ... of quantum theory since 1910 physics has gradually been compelled to abandon its proud claim that it can, in principle, understand the whole universe."

Although Bohr's views are subtle and cannot be easily relegated to any particular philosophical "pigeon hole", there is no doubt that he and Pauli rejected scientific realism (both WSR and SSR) as I have characterized it (see Murdoch, 1987, for a discussion of the realism question in Bohr's philosophy).

As an alternative, scientific anti-realism comes in as many varieties as does realism; I will simply outline two possible positions again labeled strong and weak. Strong scientific anti-realism (SSAR) is the doctrine that (1) science does not aim to provide an accurate representation of the world as it is in itself, but rather exclusively seeks empirical adequacy[12] and (2) belief in the entities postulated by scientific theories, if they are not independently verifiable (generally via direct observation) is unwarranted and irrational.[13] Weak scientific anti-realism (WSAR) embraces only clause (1) but does not strictly enjoin against belief in unobservable entities of the sort typically postulated by scientific theorizing. Bohr probably regarded the claim that nature outruns the limits of human observability as too obvious to bear comment, and the relevant scientific experiments – under their *classical* descriptions – were designed to open a gateway to the realm of the invisibly small. However, it does not follow from this concession that science is in any way revealing micronature as it is in itself. It is thus fair to say that Bohr and Pauli embraced WSAR rather than SSAR.

---

[12] Empirical adequacy is the ability of theories to predict accurately the results of experiments or observations. It comes in degrees and the more a theory possesses the better the theory. There is no rational inference – according to the anti-realist – from empirical adequacy to the truth of the theory.

[13] Like the extreme form of realism – SSR – it is doubtful that anyone holds SSAR as I have defined it. The closest position is that of Bas van Fraassen, whose radical anti-realism only balks at the claim that belief in unobservable entities is irrational. But that is because van Fraassen has a very liberal notion of rationality in which any view counts as potentially rational so long as it is not ruled out by the dictates of logic or mathematics (especially the principles of probability theory) or has been demonstrated to fail the test of empirical adequacy (Fraassen, 2002).

6 Complementarity as a Basis of Weak Anti-Realism

But why did Pauli reject scientific realism ? I think the fundamental reason is the revelation of complementarity which can be understood as flatly asserting that it is impossible for there to be a complete scientific description of nature. Even within a strictly scientific domain dealing with properties which one would expect that only science could get a grip of, e.g. the quantum analogues of classical properties such as momentum or position and also purely quantum mechanical properties such as spin, isospin, quark colors, etc., no single complete description is possible. Instead, we find that there are many incompatible ways of describing nature, and it is – to a point – up to the experimenter which description shall be employed in any given experimental setup. The analogy with a Spinozistic view of one "substance" which is capable of sustaining multiple essentially distinct characterizations is obvious here, even though the so-called Copenhagen interpretation of quantum mechanics does not demand that we extend the idea of complementarity beyond the scientific domain (although that is exactly the strategy which Pauli followed).

The drive towards anti-realism can be exhibited if we contrast Einstein's attitude with that of Pauli about this essential "incompleteability" of a purely scientific description of nature. As is well known, Einstein regarded this as a fundamental defect in quantum mechanics which, he was perennially inclined to think, led to actual paradox, but at least to philosophically unacceptable consequences. Pauli (in Meier, 2001, p. 121) took the opposite view, saying that "... Einstein was regarding as an imperfection of wave mechanics within physics what was an imperfection of physics within life."

This "imperfection of physics" might be called experimental relativity, according to which reality (as revealed in scientific experimentation) is conditioned by the choice of an experimental target. That is, physical reality can only be apprehended in partial and mutually exclusive (complementary) conceptual schemes and there is no way to account for, or factor out, the effects of choosing one scheme in an experimental setup among complementary schemes. Hence, there is no way to describe physical reality apart from the experimental setups which probe it. Furthermore, this basic constraint on scientific description transcends the usual modes of operation of typical scientific laws: it is non-local, non-causal, non-energetic. It seems Pauli took this for a sign that physical science was not revealing new features of scientifically describable reality but rather butting up against the limitations of science itself. Science is revealed to be not the ultimate guide to reality but merely a limited way of characterizing something more fundamental (Pauli, 1950, p. 78; in Enz and Meyenn, 1994, p. 40):

> "... The precondition for a description of phenomena independently of the mode of their observation is no longer fulfilled, and physical objects acquire a two-valued, or many-valued, and therefore symbolic character."

This is an interesting argument for anti-realism in which we find the realist position to be self-defeating. That is, the assumption that science provides the truth about the ultimate nature of reality leads – via the historical development of quantum mechanics – to the "paradox" of experimental relativity. And experimental relativity leads us to deny that science can reveal nature as it is in itself.

However, Pauli did *not* mean to say by this that physics is incomplete in the sense that some factor involved in the evolution of the physical world is left out of the quantum mechanical description which a more perfect physical description could, in principle, incorporate and thus retrieve the metaphysical pretension to describe the world in its totality as it is in itself. Nor is there an extra-physical dimension to the world which acts in causal commerce with its physical evolution. In particular, there is no intrusion of psyche into physical workings – physics is fully objective. In contrast to some interpretations of quantum mechanics that assign a distinguished, quasi-dynamical role to the mind of the observer (or consciousness in general), Pauli regarded the physical description as complete so far as it went. Pauli (1957, p. 44; in Enz and Meyenn, 1994, p. 133) maintained that experimental

> "results present themselves to the observers as objective reality . . . Subjective or psychical properties of the observer do not enter into the physical descriptions of nature in quantum mechanics."

And "measurement results cannot be influenced by the observer, once he has selected his experimental setup" (Pauli, 1954, p. 286; in Enz and Meyenn, 1994, p. 152).

We can extract from the foregoing reflections two core theses of Pauli's. The first is that physics is (potentially) complete and objective "in itself", while the second is that science does not provide or even aim to provide a representation of the world as it is in itself. There is some evident tension between these two theses. The former seems to support or perhaps even lead to some kind of scientific realism. If our scientific description of the world is complete, why not endorse it as our best picture of the nature of reality ? On the other hand, insofar as we deny scientific realism we seem driven to admit that the scientific picture of the world must be incomplete after all. Of course, the solution is to opt for the idea that the scientific picture of the world is *metaphysically* incomplete. The idea of complementarity serves us well again, this time affording the conceptual tool to understand how this could be.

7 Conclusions

Recall Pauli's application of the concept of complementarity to the problem of attaining knowledge of the unconscious mind. There, Pauli claimed that the psyche was a whole which could be regarded from two complementary perspectives: that of conscious introspective knowledge and indirect theoretical knowledge. The quantum analogy can be amplified. The world itself can be the whole of which we can attain knowledge only via incompatible, complementary perspectives: the mental and the physical. We cannot forbear deploying one or the other of these perspectives, and each is a legitimate source of knowledge. As Pauli (in Meier, 2001, p. 87) says:

> "It is true that in the empirical world of phenomena there must always be the difference between 'physical' and 'psychic' ... but now that matter has become an invisible reality for the modern physicist, the prospects for a psycho-physical monism have become much more favorable."

But neither perspective provides an entry into reality as it is in itself – this is the "super unknown" or what Pauli sometimes calls irrepresentable (*unanschaulich*). I think that the impossibility of representing reality as it is explains why Pauli refers to what lies "behind" the physical and psychical aspects as an "abstract fact". We can of course *refer* to it and in this minimal way also represent it, but we have no way of representing it as it is in itself – in this sense it remains an abstract idea. Occasionally Pauli even calls for the invention of a neutral language, what he calls a "psycho-physical standard language" which would (somehow) serve to describe an "invisible, potential form of reality that is only indirectly inferable ... " (Meier, 2001, p. 82).[14]

There remains to consider Pauli's remark that the mental and the physical should appear as mirror images of each other as "reflected" in the fundamental, neutral stuff of ultimate reality. This goes beyond the already audacious idea of a neutral substrate underlying both the physical and psychical aspects of reality. Nonetheless, the internal completeness of the physical picture coupled with the independence of the mental pole goes some way to suggest that both aspects ought to be complete and hence should stand as mirrors of each other.

I would like to conclude with a final issue closely connected to the problem of how mind and matter could be mirror images of each other. I can address it only briefly because, so far as I know, Pauli says almost nothing about it. This is the question of whether a dual-aspect view such as Pauli's which endorses the condition that matter and mind should be mirror images of each other entails some form of panpsychism. This is certainly the inference that Spinoza drew from his version of dual-aspect theory. According to Spinoza (1677/1985, Prop. 7, scholium), every physical entity has a correspondent in the realm of the mental, and vice versa:

---

[14] It is interesting to contrast Pauli's linguistic dream with that of Thomas Nagel (1974) who in his famous "What Is It Like to Be a Bat ?" called for a language in which the subjective could – somehow – be given an objective description. Nagel seemed to envision a way to enfold the subjective features of consciousness into a language which could integrate with physical science, whereas Pauli appears to want a language that would transcend both the mental and the physical perspective. The prospects for either project appear rather dim inasmuch as it is difficult to see how we could even begin to construct any non-trivial form of either language.

> "A circle existing in nature and the idea of the existing circle, which is also in God, are one and the same thing ... therefore, whether we conceive nature under the attribute of Extension, or under the attribute of Thought ... we shall find one and the same order, or one and the same connection of causes ... ."

In the abstract, it seems difficult to put forth the idea that mind and matter are like mirror images of each other and simultaneously deny that *every* physical entity has its mental counterpart. It must thus be acknowledged that the sort of dual-aspect theory which Pauli outlines at the very least strongly tends towards the admission of some kind of panpsychism. Unfortunately, Pauli says next to nothing explicitly about panpsychism. He does make some rather cryptic remarks in a footnote about the panpsychism of Bernhard Rensch which Rensch called "hylopsychism" that can be taken as not unfriendly to the doctrine (Pauli, 1954, p. 289; in Enz and Meyenn, 1994, p. 155). And scattered here and there in Pauli's writings is the idea that the mental realm is analogous to a physical field. Since it is a fundamental feature of such fields that they have a value at every point of space the analogy would also seem to point towards a panpsychist understand of mind's place in nature.

It would be very interesting to find out more about Pauli's view on mind and it is to be hoped that further study of his voluminous correspondence will reveal more details of the content of Pauli's dual-aspect theory of the mind-matter relation, more on his attitude towards panpsychism and further hints on the way complementarity and its attendant scientific anti-realism led Pauli to embrace a radical metaphysical viewpoint so distant from the currently dominant materialist paradigm.

References

Adams, R.M. (1994): *Leibniz: Determinist, Theist, Idealist.* Oxford University Press, Oxford.

Atmanspacher, H. and Primas, H. (2006): Pauli's ideas on mind and matter in the context of contemporary science. *Journal of Consciousness Studies* **13**(3), 5–50.

Bohr, N. (1934): *Atomic Theory and the Description of Nature.* Cambridge University Press, Cambridge.

Broad, C.D. (1925): *The Mind and Its Place in Nature.* Harcourt, Brace and Company, New York.

Cartwright, N. (1983): *How the Laws of Physics Lie.* Clarendon Press, Oxford.

Einstein, A. (1953): Einleitende Bemerkung über Grundbegriffe. In: George, A. (ed.), *Louis de Broglie. Physicien et Penseur.* Editions Albin Michel, Paris, pp. 5–15.

Enz, C.P. and Meyenn, K. von (eds.) (1994): *Wolfgang Pauli. Writings on Physics and Philosophy.* Springer, Berlin.

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

Fraassen, B.C. van (2002): *The Empirical Stance.* Yale University Press, New Haven.

Giere, R. (1988): *Explaining Science: A Cognitive Approach.* University of Chicago Press, Chicago.

Hacking, I. (1983): *Representing and Intervening.* Cambridge University Press, Cambridge.

Harré, R. (1986): *Varieties of Realism.* Blackwell, Oxford.

Meier, C.A. (ed.) (2001): *Atom and Archetype. The Pauli/Jung Letters, 1932–1958.* Princeton University Press, Princeton.

Murdoch, D. (1987): *Niels Bohr's Philosophy of Physics.* Cambridge University Press, Cambridge.

Nagel, T. (1974): What is it like to be a bat ? *The Philosophical Review* **83**, 435–450.

Pauli, W. (1950): Die philosophische Bedeutung der Komplementarität. *Experientia* **6**, 72–81.

Pauli, W. (1952): Der Einfluss archetypischer Vorstellungen auf die Bildung naturwissenschaftlicher Theorien bei Kepler. In: Jung, C.G., and Pauli, W. (eds.), *Naturerklärung und Psyche.* Rascher, Zürich, pp. 109–194.

Pauli, W. (1954): Naturwissenschaftliche und erkenntnistheoretische Aspekte der Ideen vom Unbewussten. *Dialectica* **8**, 283–301.

Pauli, W. (1956): Die Wissenschaft und das abendländische Denken. In: Göhring, M. (ed.), *Europa – Erbe und Aufgabe. Internationaler Gelehrtenkongress, Mainz 1955.* Steiner, Wiesbaden, pp. 71–79.

Pauli, W. (1957): Phänomen und physikalische Realität. *Dialectica* **11**, 36–48.

Petersen, A. (1963): The philosophy of Niels Bohr. *Bulletin of the Atomic Scientist* **19**(7), 8–14.

Place, U.T. (1956): Is consciousness a brain process ? *British Journal of Psychology* **47**, 44–50.

Reichenbach, H. (1938): *Experience and Prediction.* University of Chicago Press, Chicago.

Ryle, G. (1949): *The Concept of Mind.* Hutchinson, London.

Schmaltz, T. (2006): Nicolas Malebranche. *Stanford Encyclopedia of Philosophy,* `plato.stanford.edu/archives/sum2006/entries/malebranche/`.

Seager, W. (2007): A brief history of the philosophical problem of consciousness. In: Zelazo, P.D., Moscovitsh, M., and Thompson, E. (eds.), *The Cambridge Handbook of Consciousness.* Cambridge University Press, Cambridge, pp. 9–33.

Shapiro, L. (2007): *The Correspondence between Princess Elisabeth of Bohemia and René Descartes.* University of Chicago Press, Chicago.

Smart, J.J.C. (1959): Sensations and brain processes. *The Philosophical Review* **68**, 141–156.

Spinoza, B (1677): Ethics. In: Curley, E.M. (ed.), *The Collected Works of Spinoza. Volume I.* Princeton University Press, Princeton 1985.

Stubenberg, L. (2005): Neutral monism. *Stanford Encyclopedia of Philosophy,* `plato.stanford.edu/archives/spr2005/entries/neutral-monism/`.

Thagard, P. (1992): *Conceptual Revolutions.* Princeton University Press, Princeton.

Extending the Philosophical Significance of the Idea of Complementarity

Peter beim Graben[1] and Harald Atmanspacher[2]

[1] School of Psychology and Clinical Language Sciences, University of Reading, UK, `p.r.beimgraben@reading.ac.uk`
[2] Institute for Frontier Areas of Psychology, Freiburg, Germany, `haa@igpp.de`

Summary. We discuss a specific way in which the notion of complementarity can be based on the dynamics of the system considered. This approach rests on an epistemic representation of system states, reflecting our knowledge about a system in terms of coarse grainings (partitions) of its phase space. Within such an epistemic quantization of classical systems, compatible, comparable, commensurable, and complementary descriptions can be precisely characterized and distinguished from each other. Some tentative examples are indicated that, we suppose, would have been of interest to Pauli.

1 Introduction

In 1949 Pauli delivered a lecture on complementarity to the Philosophical Society in Zurich, which was then published (in German) under the title "The Philosophical Significance of the Idea of Complementarity" in the journal *Experientia* (Pauli, 1950). His article followed an earlier paper by Bernays (1948) "On the Extension of the Notion of Complementarity into Philosophy" (also in German). Pauli (1950) emphasized that the

> "situation in regard to complementarity within physics leads naturally beyond the narrow field of physics to analogous situations in connection with the general conditions of human knowledge."

Pauli's paper addressed a number of pertinent topics still unresolved today where the idea of complementarity might be of relevance, such as "the experimenter's free choice between mutually exclusive experimental arrangements", "the idea of the cut between observer or instrument of observation and the system observed" (nowadays dubbed the *Heisenberg cut*), "considerations of purposefulness" concerning the actual location of the cut, and eventually the "paradoxical" relationship between consciousness and the unconscious (Pauli, 1950):

> "On one hand, modern psychology demonstrates a largely objective reality of the unconscious psyche; on the other hand every bringing into consciousness, i.e. observation, constitutes an interference with the unconscious contents that is in principle uncontrollable; this limits the objective character of the reality of the unconscious and invests reality with a certain subjectivity."

The concept of complementarity was introduced into physics by Bohr (cf. Bohr, 1948), but he was familiar with it from psychological texts by William James and from his psychologist friend Arthur Rubin, who studied the perception of ambiguous stimuli. In simple words, two descriptions of a situation are complementary if they exclude each other and yet are both necessary for an exhaustive description of that situation. In quantum theory, this vague characterization was made much more precise in the mathematical framework of non-commutative algebras or non-Boolean lattices of quantum observables. The price to be paid for this precision is the restriction of the concept of complementarity to quantum physics.

However, there are many more candidates for complementary relationships in other sciences, e.g. in psychology and philosophy. The present article intends to reconsider the foundations of the notion of complementarity not only with respect to quantum systems but with a broader domain of applications.[3] It builds essentially on a recent paper by beim Graben and Atmanspacher (2006) which describes in technical detail how complementary observables can be defined in classical physical systems if their dynamics is taken into account properly. In the present paper we give a simplified exposition for a more general readership and address some issues that were in the focus of Pauli's interest for many years.

Section 2 contains a compact reminder of how complementarity and compatibility are defined in quantum theory. Section 3 introduces the concept of partitions for an epistemic treatment of classical dynamical systems. Section 4 illustrates how complementary observables can be introduced for epistemic states defined on the basis of particular phase space partitions. Only if such partitions are generating (or, more specifically, Markov), they define epistemic states that are stable under the dynamics and provide compatible epistemic descriptions. Partitions chosen more or less *ad hoc* generally lead to incompatible or complementary descriptions. Section 5 characterizes and delineates compatible, comparable, and commensurable theories (and their opposites) from each other. Some examples are outlined in Sect. 6.

---

[3] Compare Atmanspacher *et al.* (2002) and Primas (2007) for formally rigorous approaches in this direction.

2 Compatibility and Complementarity in Quantum Theory

In quantum theory, measurements of observables A and B with pure point spectra which produce dispersion-free values as results depend in general on the sequence in which the measurements are carried out. In this case, the observables A, B are called *incompatible*. If, on the other hand, the order of measuring A and B does not play a role, the observables A, B are called *compatible*. Therefore, compatibility can be formally expressed by the equation

$$AB \;=\; BA, \tag{1}$$

while incompatibility means that A and B do not commute:

$$AB \;\neq\; BA. \tag{2}$$

In a Hilbert space representation, (1) has the consequence that compatible observables are simultaneously diagonalizable, i.e. all eigenstates of A are also eigenstates of B (and vice versa), and these common eigenstates span the whole Hilbert space of (pure) quantum states. Since $A\psi = a\psi$ for eigenstates ψ with eigenvalue a of A, observable A assumes the sharp, *dispersion-free* value a in eigenstate ψ.

Compatible observables with pure point spectra are therefore dispersion-free in their common eigenstates which span the whole Hilbert space. Incompatible observables do not share all eigenstates, although they may share some of them. *Complementary* observables can be characterized as being *maximally incompatible*; they do not have any eigenstate in common (beim Graben and Atmanspacher, 2006). These results will be used for generalizing the concepts of complementarity and compatibility to classical systems, i.e. beyond quantum systems, in the next sections.

3 Epistemic Descriptions of Classical Dynamical Systems

Measurements (or observations) require the preparation of a state of the system to be measured (or observed), choices of initial and boundary conditions for this state, and the selection of particular measurement setups. They refer to operationally defined *observables* which can be deliberately chosen by the experimenter (Pauli, 1950; Primas, 2007).

3.1 Observables and Partitions

A classical dynamical system is characterized by the fact that all observables are compatible with each other. However, in general this holds only for a so-called *ontic description* (Atmanspacher, 2000) where the state of a system is considered as if it could be characterized precisely as it is (relative to

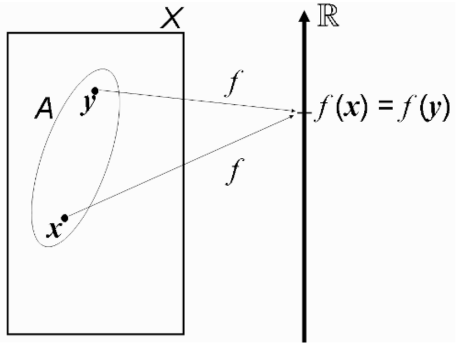


Fig. 1. States $\boldsymbol{x}, \boldsymbol{y}$ in a phase space X of a classical system (left) and the real numbers as the range of a classical observable $f : X \to \mathbb{R}$ (right). Epistemically equivalent states $\boldsymbol{x}, \boldsymbol{y} \in X$ belong to the same equivalence class $A \subset X$.

a chosen ontology (Quine, 1969; Atmanspacher and Kronz, 1999; Dale and Spivey, 2005). On such an account, the *ontic state* of the system is given by a point $\boldsymbol{x}$ in phase space X. The associated observables are real-valued functions $f : X \to \mathbb{R}$, such that $a = f(\boldsymbol{x})$ is the value of f in state $\boldsymbol{x}$. By contrast, *epistemic descriptions* refer to the "knowledge that can be obtained about an ontic state" (Atmanspacher, 2000). For the sake of simplicity we shall identify *epistemic states* with subsets $S \subset X$ in phase space, thus expressing that they can be specified only with limited accuracy.

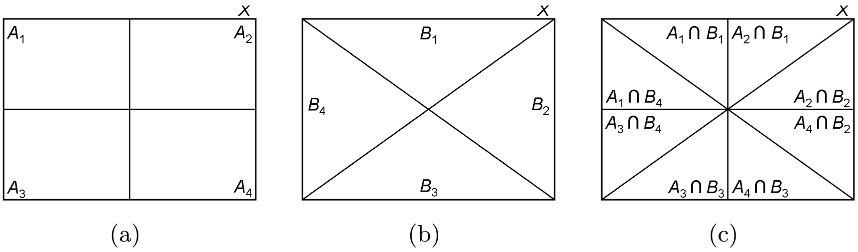
Figure 1 displays a situation in which the observable f is not injective, such that different states $\boldsymbol{x} \neq \boldsymbol{y} \in A \subset X$ lead to the same measurement result

$$f(\boldsymbol{x}) = f(\boldsymbol{y}). \tag{3}$$

In this case, the states $\boldsymbol{x}$ and $\boldsymbol{y}$ are *epistemically* indistinguishable by means of the observable f (Shalizi and Moore, 2003; beim Graben and Atmanspacher, 2006). Measuring f cannot tell us whether the system is in state $\boldsymbol{x}$ or $\boldsymbol{y}$. The two states are therefore *epistemically equivalent* with respect to f (beim Graben and Atmanspacher, 2006).

In this way, the observable f induces an equivalence relation "$\sim_f$" on the phase space X: $\boldsymbol{x} \sim_f \boldsymbol{y}$ if $f(\boldsymbol{x}) = f(\boldsymbol{y})$. The resulting equivalence classes of ontic states partition the phase space into mutually exclusive and jointly exhaustive sets $A_1, A_2, \ldots$ such that $A_i \cap A_j = \emptyset$ for all $i \neq j$ and $\bigcup_i A_i = X$. These sets are the epistemic states that are induced by the observable f. The collection $\mathcal{F} = \{A_1, A_2, \ldots\}$ of epistemic states is a phase space *partition.*

We call f an *epistemic observable* if the partition $\mathcal{F}$ is not the *identity partition* $\mathcal{I}$ where every cell A_k is a singleton set containing exactly one element $A_k = \{\boldsymbol{x}_k\}$ (Shalizi and Moore, 2003). In this limiting case, f is injective and can be called an *ontic observable.* In the opposite limit, epistemic observables are constant over the whole phase space: $f(\boldsymbol{x}) = \text{const}$ for all $\boldsymbol{x} \in X$. In this



(a) (b) (c)

Fig. 2. Examples for finite partitions of the phase space X: (a) "rectangular" partition $\mathcal{F} = \{A_1, A_2, A_3, A_4\}$, (b) "triangular" partition $\mathcal{G} = \{B_1, B_2, B_3, B_4\}$, (c) product partition $\mathcal{F} \vee \mathcal{G}$.

case, all states are epistemically equivalent with each other and belong to the (same) equivalence class X of the *trivial partition* $\mathcal{T}$.

Most interesting for our purposes are finite partitions $\mathcal{F} = \{A_1, A_2, \dots A_n\}$ (where n is a finite natural number) which are neither trivial nor identity. Figures 2(a,b) display two different finite partitions. From the partitions $\mathcal{F}$ and $\mathcal{G}$ shown in Figs. 2(a,b), a *product partition*, $\mathcal{P} = \mathcal{F} \vee \mathcal{G}$ can be constructed. This partition, depicted in Fig. 2(c), contains all possible intersections of sets in $\mathcal{F}$ with sets in $\mathcal{G}$:

$$\mathcal{P} \;=\; \mathcal{F} \vee \mathcal{G} \;=\; \{A_i \cap B_j | A_i \in \mathcal{F}, B_j \in \mathcal{G}\}. \qquad (4)$$

The product partition $\mathcal{P}$ is a *refinement* of both partitions $\mathcal{F}$ and $\mathcal{G}$. The refinement relation introduces a partial ordering relation "$\prec$" among partitions. If $\mathcal{G}$ is a refinement of $\mathcal{F}$, $\mathcal{G} \prec \mathcal{F}$, then there is a "factor partition" $\mathcal{H}$ such that $\mathcal{G} = \mathcal{F} \vee \mathcal{H}$. If neither $\mathcal{G}$ is a refinement of $\mathcal{F}$ nor *vice versa* (and $\mathcal{G} \neq \mathcal{F}$), the partitions $\mathcal{G}$ and $\mathcal{F}$ have been called *incomparable* (Shalizi and Moore, 2003).

3.2 Dynamics

A dynamical system evolves as a function of parameter time t. In other words, any present state (e.g. an initial condition) in phase space, $\boldsymbol{x}_0 \in X$, gives rise to future states $\boldsymbol{x}_t \in X$. This evolution is described by a flow map $\Phi : X \to X$. In the simple case of a deterministic dynamics in discrete time, Φ maps any state $\boldsymbol{x}_t$ onto a state $\boldsymbol{x}_{t+1}$, as illustrated in Fig. 3. Iterating the map Φ, yields a *trajectory* of states

$$\boldsymbol{x}_{t+1} \;=\; \Phi^{t+1}(\boldsymbol{x}_0) \;=\; \Phi(\Phi^t(\boldsymbol{x}_0)) \;=\; \Phi(\boldsymbol{x}_t) \qquad (5)$$

for integer positive times $t \in \mathbb{N}$. Likewise, the inverse map Φ^{-1} can be iterated if the dynamics is invertible: $\boldsymbol{x}_{-(t+1)} = \Phi^{-(t+1)}(\boldsymbol{x}_0) = \Phi^{-1}(\Phi^{-t}(\boldsymbol{x}_0)) = \Phi^{-1}(\boldsymbol{x}_{-t})$, again for integer positive times $t \in \mathbb{N}$. Therefore, the dynamics of an invertible discrete-time system is described by the one-parameter group of integer numbers $t \in \mathbb{Z}$.

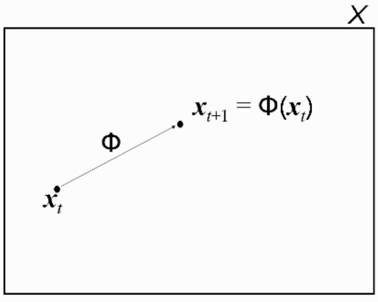


Fig. 3. A discrete-time dynamics of a classical system is given by a map $\Phi : X \to X$ which assigns to a state $\boldsymbol{x}_t$ at time t its successor $\boldsymbol{x}_{t+1} = \Phi(\boldsymbol{x}_t)$ at time $t + 1$.

3.3 Continuous Measurements

In Sect. 3.1, we have described instantaneous measurements by the action of an observable $f : X \to \mathbb{R}$ on an ontic state $\boldsymbol{x}$. Now we are able to describe *continuous measurements*[4] by combining the action of an observable f with the dynamics Φ. Let the system be in state $\boldsymbol{x}_0 \in X$ at time $t = 0$. Measuring $f(\boldsymbol{x}_0)$ tells us to which class of epistemically equivalent states in the partition $\mathcal{F}$, associated with f, the state $\boldsymbol{x}_0$ belongs. Suppose that this is the cell $A_{i_0} \in \mathcal{F}$. Suppose further that measuring f in the subsequent state $\boldsymbol{x}_1 = \Phi(\boldsymbol{x}_0) \in X$ reveals that $\boldsymbol{x}_1$ is contained in another cell $A_{i_1} \in \mathcal{F}$.

An alternative way to describe this situation is to say that the initial state $\boldsymbol{x}_0 = \Phi^{-1}(\boldsymbol{x}_1)$ belongs to the pre-image $\Phi^{-1}(A_{i_1})$ of A_{i_1}. The information about $\boldsymbol{x}_0$ that is gained by measuring $f(\boldsymbol{x}_1)$ is, then, that the initial state $\boldsymbol{x}_0$ was contained in the intersection $A_{i_0} \cap \Phi^{-1}(A_{i_1})$. Continuing the observation of the system over one more instant in time yields that the initial state $\boldsymbol{x}_0$ belonged to the set $A_{i_0} \cap \Phi^{-1}(A_{i_1}) \cap \Phi^{-2}(A_{i_2})$ if the third measurement result was $\boldsymbol{x}_2 = \Phi^2(\boldsymbol{x}_0) \in A_{i_2}$.

A systematic investigation of continuous measurements relies on the definition of the pre-image of a partition,

$$\Phi^{-1}(\mathcal{F}) \; = \; \{\Phi^{-1}(A_i) | A_i \in \mathcal{F}\}, \tag{6}$$

which consists of all pre-images of the cells A_i of the partition $\mathcal{F}$. Then, a continuous measurement over two successive time steps is defined by the product partition $\mathcal{F} \vee \Phi^{-1}(\mathcal{F})$, containing all intersections of cells of the original partition $\mathcal{F}$ with cells of its pre-image $\Phi^{-1}(\mathcal{F})$. The result of the measurement of f over two time steps is $\boldsymbol{x}_0 \in A_{i_0} \cap \Phi^{-1}(A_{i_1}) \subset \mathcal{F} \vee \Phi^{-1}(\mathcal{F})$. This product partition is called the *dynamic refinement* of $\mathcal{F}$, illustrated in Fig. 4.

---

[4] The notion of a continuous measurement does not refer to continuous time $t \in \mathbb{R}$ but characterizes that a measurement extends over time. This can also be the case for discrete time $t \in \mathbb{Z}$.

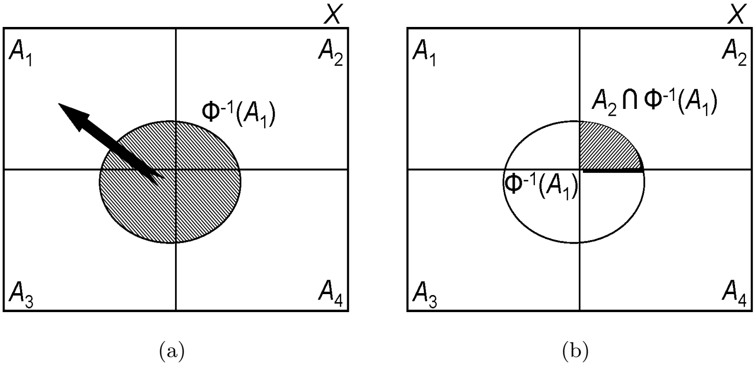


Fig. 4. Dynamic refinement of a partition. (a) For each cell A_i of the partition $\mathcal{F}$ the pre-image $\Phi^{-1}(A_i)$ under the dynamics is determined. The bold arrow indicates that the shaded region in phase space is mapped onto cell A_1. (b) The shaded region in the product partition $\mathcal{F} \vee \Phi^{-1}(\mathcal{F})$ is the element $A_2 \cap \Phi^{-1}(A_1)$ of the dynamically refined partition.

Most information about the state of a system can be gained by an ideal, "ever-lasting" continuous measurement that began in the infinite past and terminates in the infinite future. This leads to the *finest dynamic refinement*

$$\mathbf{R}\mathcal{F} = \bigvee_{t=-\infty}^{\infty} \Phi^t(\mathcal{F}), \tag{7}$$

expressed by the action of the "finest-refinement operator" $\mathbf{R}$ upon a partition $\mathcal{F}$. It would be desirable that such an ever-lasting measurement yields complete information about the initial condition $\boldsymbol{x}_0$ in phase space. This is achieved if the refinement (7) entails the identity partition,

$$\mathbf{R}\mathcal{F} = \mathcal{I}. \tag{8}$$

A partition $\mathcal{F}$ obeying (8) is called *generating.*

Given the ideal finest refinement $\mathbf{R}\mathcal{F} = \mathcal{P}$ of a (generating or non-generating) partition $\mathcal{F}$ that is induced by an epistemic observable f, we are able to regain a description of continuous measurements of arbitrary finite duration by joining subsets of $\mathcal{P}$ which are visited by the system's trajectory during measurement. Supplementing the "join" operation by the other Boolean set operations over $\mathcal{P}$ leads to a *partition algebra* $A(\mathcal{P})$ of $\mathcal{P}$. Then, every set in $A(\mathcal{P})$ is an epistemic state measurable by f.

Note that the concept of a generating partition in the ergodic theory of deterministic systems is related to the concept of a *Markov chain* in the theory of stochastic systems. Every deterministic system of first order gives rise to a Markov chain which is generally neither ergodic nor irreducible. Such Markov chains can be obtained by so-called *Markov partitions* that exist for expanding or hyperbolic dynamical systems (Sinai, 1968; Bowen, 1970; Ruelle, 1989). For non-hyperbolic systems no corresponding existence theorem is available, and the construction can be even more tedious than for hyperbolic systems (Viana *et al.*, 2003). For instance, both Markov and generating partitions for nonlinear systems are generally non-homogeneous. In contrast to Figure 2, their cells are typically of different size and form.

Note further that every Markov partition is generating, but the converse is not necessarily true (Crutchfield, 1983; Crutchfield and Packard, 1983). For the construction of "optimal" partitions from empirical data it is often more convenient to approximate them by Markov partitions (Froyland, 2001).

4 Compatibility and Complementarity in Classical Dynamical Systems

If a partition $\mathcal{F}$ is not generating, its finest refinement is not the identity partition. In this case, the refinement operator produces a partition $\mathcal{P} = \mathbf{R}\mathcal{F}$ with some residual coarse grain. Moreover, the cells of a non-generating partition are not stable under the dynamics Φ, so that they become dynamically ill-defined – a disaster for any attempt to formulate a properly robust coarse-grained description (Atmanspacher and beim Graben, 2007).

Let $P \in \mathcal{P}$ be an epistemic state of the finest refinement of $\mathcal{F}$. Because $\mathcal{F}$ is induced by an observable f whose epistemic equivalence classes are the cells of $\mathcal{F}$, all cells of $\mathcal{P}$ can be accessed by continuous measurements of f. However, as $\mathcal{P}$ is not the identity partition $\mathcal{I}$, the singleton sets $\{\boldsymbol{x}\}$ representing ontic states in X are not accessible by measuring f. An arbitrary epistemic state $S \subset X$ induced by an observable g is called *epistemically accessible with respect to* f (beim Graben and Atmanspacher, 2006) if S belongs to the partition algebra $A(\mathcal{P})$ produced by the finest refinement of $\mathcal{F}$.

Measuring the observable f in all ontic states $\boldsymbol{x} \in P$ belonging to an epistemic state $P \in \mathcal{P}$ always yields the same result $a = f(\boldsymbol{x})$ since f is by construction constant over P. Therefore, the variance of $f(\boldsymbol{x})$ across P vanishes such that f is dispersion-free in the epistemic state P. In other words, P is an eigenstate of f. One can now easily construct another observable g that is not dispersion-free in P such that P is not a common eigenstate of f and g. According to Sect. 2, the observables f and g are, thus, incompatible as they do not share all (epistemically accessible) eigenstates. Beim Graben and Atmanspacher (2006) refer to this construction as an *epistemic quantization* of a classical dynamical system.

In an ontic description of a classical system, ontic states are common eigenstates of all observables. Therefore, classical observables associated with ontic states are always compatible. By contrast, if the ontic states are not epistemically accessible by continuous measurements, the smallest epistemically accessible states are cells in the finest refinement of a partition $\mathcal{F}$ induced by an epistemic observable f. These epistemic states are not eigenstates of every observable, such that observables associated with them are incompatible. As in quantum theory, two observables f and g are complementary if they do not have any (epistemically accessible) eigenstate in common, i.e. if they are maximally incompatible.

Nevertheless, even in an epistemic description, classical observables f and g can be compatible with each other. This is the case if all ontic states $\boldsymbol{x} \in X$ are epistemically accessible with respect to both f and g. The necessary and sufficient condition for this is that the partitions $\mathcal{F}$, $\mathcal{G}$ be generating (Eq. 8). This leads to a generalization of the concepts of compatibility and complementarity: Two partitions $\mathcal{F}, \mathcal{G}$ are called compatible if and only if they are both generating: $\mathbf{R}\mathcal{F} = \mathbf{R}\mathcal{G} = \mathcal{I}$. They are incompatible if $\mathbf{R}\mathcal{F} \neq \mathbf{R}\mathcal{G}$, which is always the case if at least one partition is not generating. They are complementary if their finest refinements are disjoint: $\mathbf{R}\mathcal{F} \cap \mathbf{R}\mathcal{G} = \emptyset$.[5]

These definitions give rise to three main corrolaries. (1) For compatible partitions, every ontic state $\boldsymbol{x}$ is epistemically accessible with respect to observables f, g inducing the partitions $\mathcal{F}$, $\mathcal{G}$. Hence, every ontic state is a common eigenstate of f and g and all ontic states span the whole phase space $X = \bigcup_{\boldsymbol{x}} \{\boldsymbol{x}\}$. (2) For incompatible partitions, epistemically accessible eigenstates of one observable are not necessarily epistemically accessible eigenstates of another observable. (3) For complementary partitions, the observables do not have any eigenstates in common and are therefore maximally incompatible.

5 Compatible, Comparable, and Commensurable Theories

A proposition such as “the observable f assumes the value a in state $\boldsymbol{x} \in X$”, or briefly “$a = f(\boldsymbol{x})$”, induces a binary partition of the phase space X of a classical dynamical system into two subsets,

$$\mathcal{F} = \{S, X \setminus S\}, \tag{9}$$

where $S = \{\boldsymbol{x} \in X | a = f(\boldsymbol{x})\}$. Because propositions can be combined by the logical connectives “and”, “or”, and “not”, the structure of a *classical theory* is that of a Boolean algebra of subsets of the phase space (Primas, 1977; Westmoreland and Schumacher, 1993; Primas, 2007). In the following we shall elucidate such theories with respect to the epistemic quantization discussed in Sect. 4.

Given a classical dynamical system with phase space X, dynamics Φ, and a family of appropriately chosen epistemic observables $f_1, f_2, \dots f_n$, these observables induce partitions $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$ whose product $\mathcal{F} = \bigvee_{i=1}^{n} \mathcal{F}_i$ characterizes one particular setup for possible measurements. The partition algebra

---

[5] These concepts can also be defined by means of σ-algebras in measure theory (beim Graben and Atmanspacher, 2006). For the present simplified exposition, which captures very much the same idea, set-theoretical concepts are sufficient.

$A(\mathcal{F})$, comprising all subsets of X that can be formed by the Boolean set operations "join", "intersection", and "difference", can be identified with a classical theory of propositions corresponding to instantaneous measurements of one of the observables $f_1, f_2, \dots f_n$ on X.

For continuous measurements, the dynamic refinement according to (7) has to be taken into account. In this case we have to consider the partition algebra $A(\mathbf{R}\mathcal{F})$ in order to form propositions about continuous measurements of arbitrary duration. Hence, a *classical theory* $T(\mathcal{F})$ refers to the Boolean set algebra $A(\mathbf{R}\mathcal{F})$ over the finest refinement $\mathbf{R}\mathcal{F}$.

Using the results of Sect. 4, two theories $T(\mathcal{F})$ and $T(\mathcal{G})$ are called compatible if their partitions $\mathcal{F}$ and $\mathcal{G}$ are compatible (i.e. if $\mathcal{F}$ and $\mathcal{G}$ are both generating). They are called incompatible if their partitions are incompatible, and they are called complementary if their partitions are complementary. The "experimenter's free choice between mutually exclusive experimental arrangements" (Pauli, 1950; Primas, 2007) corresponds to the choice of incompatible or complementary theories that are based upon non-generating partitions associated to epistemic observables. Insofar as classical ontic observables always induce the identity partition $\mathcal{I}$ on the phase space, ontic theories are always compatible with each other.

Following Shalizi and Moore (2003), we call two theories $T(\mathcal{F})$ and $T(\mathcal{G})$ *comparable* with each other if either $\mathbf{R}\mathcal{F}$ is a refinement of $\mathbf{R}\mathcal{G}$, or $\mathbf{R}\mathcal{G}$ is a refinement of $\mathbf{R}\mathcal{F}$, or $\mathbf{R}\mathcal{F} = \mathbf{R}\mathcal{G}$. Two theories are *incomparable* if they are not comparable. It is easy to realize that compatible theories are also comparable, as $\mathbf{R}\mathcal{F} = \mathbf{R}\mathcal{G} = \mathcal{I}$. However, even incompatible theories might be comparable, e.g. if one of them is based on a generating partition.

Another notion related to compatibility and comparability is that of *commensurability* (Kuhn, 1983; Hoyningen-Huene, 1990), which has gained some popularity in relativist accounts within the philosophy of science. Two theories are said to be commensurable if there is a common theoretical language that can be used to compare them. Following Primas (1977), this can be reformulated by saying that two theories $T(\mathcal{F}), T(\mathcal{G})$ are commensurable if they can be embedded into one *universal theory* $T(\mathcal{U})$ such that $T(\mathcal{F}), T(\mathcal{G})$ are sub-theories of $T(\mathcal{U})$.

More specifically, we call two theories $T(\mathcal{F}), T(\mathcal{G})$ commensurable if a theory $T(\mathcal{U})$ exists such that $\mathbf{R}\mathcal{U}$ is a refinement of $\mathbf{R}\mathcal{F}$ and $\mathbf{R}\mathcal{G}$. If $\mathcal{F}$ and $\mathcal{G}$ are both generating partitions, $\mathbf{R}\mathcal{F} = \mathbf{R}\mathcal{G} = \mathcal{I}$. Then $T(\mathcal{I})$ is a common refinement of $T(\mathcal{F})$ and $T(\mathcal{G})$, entailing that compatible theories are always commensurable. Comparable theories, whose partitions are refinements of each other, are trivially commensurable.

6 Examples

Let us finally give some selected examples for how the notions of compatibility, comparability, and complementarity can be useful for the discussion of topics within Pauli's lifelong interest.

A first illustrative example refers back to where Bohr became familiar with the notion of complementarity: the bistable perception of ambiguous stimuli. The involved processes can be described as (i) an oscillation between the two possible representations of the stimulus, and (ii) a projection into one of them, mimicking its observation. These two processes can indeed be shown to be complementary (Atmanspacher *et al.*, 2008) in basically the same sense as complementarity in quantum physics is due to non-commuting observables.

Along a different vein, beim Graben (2004) discussed three examples of implementations of symbol processors that are generically incompatible with respect to different partitions. This is due to the fact that, in these examples, the partitions are not generating. As an important consequence of this result, symbolic and subsymbolic (e.g. neural) descriptions of cognitive processes are incompatible in general. This confirms – though on different grounds – an assertion by Smolensky (1988, 2006) that an *integrated connectionist/symbolic architecture* is mandatory for cognitive science, where (Smolensky, 2006)

> "higher cognition must be formally characterized on two levels of description. At the microlevel, parallel distributed processing (PDP) characterizes mental processing; this PDP system has special organization in virtue of which it can be characterized at the macrolevel as a kind of symbolic computational system."

However, the apparent algorithmic behavior at the symbolic macrolevel is not implemented by algorithms performed at the microlevel. The microlevel dynamics only "approximates" the symbolic computations at the macrolevel, thus making both levels incompatible with each other (Smolensky, 1988). This shows any discomfort about the lack of a coherent unified framework for cognitive science to be misplaced. Incompatible descriptions are unavoidable and not an obstacle that one may hope to overcome some day. This applies also to incompatibilities and incommensurabilities in psychological theories (Yanchar and Slife, 1997; Slife, 2000; Dale and Spivey, 2005) as discussed by Atmanspacher and beim Graben (2007).

This relates to an issue raised by Pauli (1950) in terms of "considerations of purposefulness" for choosing between incompatible descriptions. An example is the notion of an *intended partition* for the dynamical systems approach to cognition (beim Graben, 2004). Among the many possible (and presumably incompatible) partitions of a dynamical system only a few, either explicitly constructed or evolutionarily optimized, give rise to a high-level interpretation of the system's low-level behavior in terms of symbol processing or cognitive computation. Such intended partitions are able to shed light onto the *symbol grounding problem* (Harnad, 1990; Atmanspacher and beim Graben, 2007).

Yet another incompatibility, maybe even complementarity, was proposed by Pauli between conscious and unconscious mental states (see Sect. 1 and Pauli, 1950). In an afterword to his essay "On the Nature of the Psyche", Jung (1969, §439, footnote 130) quotes Pauli with the statement that

> "the epistemological situation with regard to the concepts 'conscious' and 'unconscious' seems to offer a pretty close analogy to the ... situation in physics. ... From the standpoint of the psychologist, the 'observed system' would consist not of physical objects only, but would also include the unconscious, while consciousness would be assigned the role of 'observing medium'."

In other words: mental objects and their mental environments are conceived to be generated by the transformation of elements of the unconscious into consciously and, thus, epistemically accessible categories. As long as elements of the unconscious are not yet transformed into conscious categories, they remain unconscious, and whenever a category is generated and becomes consciously accessible, it leaves the domain of the unconscious. In this sense, conscious and unconscious domains are mutually disjoint, yet they are both together necessary to characterize the mental as a whole.

In such a framework of thinking, the unconscious is explicitly conceived as part of the mental. This is in contrast to many modern accounts, in which ongoing brain activity, i.e. the dynamics of subsystems of the *material brain*, is referred to by the notion of the unconscious. This brings us to the relation between mental and material states, or the mind-matter problem as the most general topic mentioned in this section. Among the many proposals that have been made to address this problem, Pauli (1952) emphasized the idea of an ontically monistic and epistemically dualistic, namely complementary, relationship between mind and matter:

> "The general problem of the relationship between psyche and physis, between inside and outside, will hardly be solved with the notion of a 'psychophysical parallelism', put forward in the past century. However, modern science has perhaps brought us closer to a more satisfying conception of this relationship insofar as it introduced the concept of *complementarity* within physics. It would be most satisfactory if physis and psyche could be conceived as complementary aspects of the same reality."

Recent publications (Walach and Römer, 2000; Atmanspacher, 2003; Römer, 2004; Primas, 2008) have tried to popularize this idea and elaborate on it.

The controversy of what constitutes the most basic aspects of reality accompanies the development of Western philosophy since its beginning. Countervailing positions favoring either *stasis*, and thus being (e.g. Parmenides), or *change*, and thus becoming (e.g. Heraclitus) followed and responded to each other time and again. It was recently shown by Römer (2006) that the corresponding distinction of *substance* and *process* can be considered as complementary in a formally anchored way.

The general scheme of thinking which such approaches follow is today called a dual-aspect or double-aspect framework, as discussed in more detail by Seager (2008). Chalmers (1995) advocated such a framework when he introduced the notion of the "hard problem of consciousness" as the problem of how to relate the first-person, phenomenal experience of a mental state to the third-person perspective characterizing the scientific (neural, cognitive, or otherwise) study of such a state. Velmans (2002, 2008) suggested to regard first-person and third-person accounts as incompatible or complementary.

Atmanspacher and beim Graben (2007) demonstrated how the *phenomenal families* introduced by Chalmers (2000), which partition the mental space of phenomenal experiences, induce a partition of the neural phase space. If this induced partition is not generating, the resulting description in terms of mental states will be incompatible with any other description. However, carefully constructed Markov partitions of the neural phase space of macroscopic brain activity, e.g. by means of EEG signals, can lead to descriptions that are compatible with mental (symbolic) descriptions.

References

Atmanspacher, H. (2000): Ontic and epistemic descriptions of chaotic systems. In Dubois, D. (ed.), *Computing Anticipatory Systems*, Springer, Berlin, pp. 465–478.

Atmanspacher, H. (2003): Mind and matter as asymptotically disjoint, inequivalent representations with broken time-reversal symmetry. *BioSystems* **68**, 19–30.

Atmanspacher, H. and Graben, P. beim (2007): Contextual emergence of mental states from neurodynamics. *Chaos and Complexity Letters* **2**, 151–168.

Atmanspacher, H., Filk, T. and Römer, H. (2008): Complementarity and bistable perception. *This volume.*

Atmanspacher, H. and Kronz, F. (1999): Relative onticity. In Atmanspacher, H., Amann, A. and Müller-Herold, U. (eds.), *On Quanta, Mind and Matter. Hans Primas in Context*, Kluwer, Dordrecht, pp. 273–294.

Atmanspacher, H., Römer, H. and Walach, H. (2002): Weak quantum theory: Complementarity and entanglement in physics and beyond. *Foundations of Physics* **32**, 379–406.

Bernays, P. (1948): Über die Ausdehnung des Begriffes der Komplementarität auf die Philosophie. *Synthese* **7**, 66–70.

Bohr, N. (1948): On the notions of causality and complementarity. *Dialectica* **2**, 312–319.

Bowen, R. (1970): Markov partitions for axiom A diffeomorphisms. *American Journal of Mathematics* **92**, 725–747.

Chalmers, D.J. (1995): Facing up to the problem of consciousness. *Journal of Consciousness Studies* **2**(3), 200–219.

Chalmers, D.J. (2000): What is a neural correlate of consciousness? In Metzinger, T. (ed.), *Neural Correlates of Consciousness*, MIT Press, Cambridge MA, pp. 17–39.

Crutchfield, J.P. (1983): *Noisy Chaos.* PhD thesis at the University of California at Santa Cruz, Sec. 5.

Crutchfield, J.P. and Packard, N.H. (1983): Symbolic dynamics of noisy chaos. *Physica D* **7**, 201–223.

Dale, R. and Spivey, M.J. (2005): From apples and oranges to symbolic dynamics: a framework for conciliating notions of cognitive representation. *Journal of Experimental and Theoretical Artificial Intelligence* **17**, 317–342.

Froyland, G. (2001): Extracting dynamical behavior via Markov models. In Mees, A.I. (ed.), *Nonlinear Dynamics and Statistics*, Birkhäuser, Boston, pp. 281–312.

Graben, P. beim (2004): Incompatible implementations of physical symbol systems. *Mind and Matter* **2**(2), 29–51.

Graben, P. beim and Atmanspacher, H. (2006): Complementarity in classical dynamical systems. *Foundations of Physics* **36**, 291–306.

Harnad, S. (1990): The symbol grounding problem. *Physica D* **43**, 335–346.

Hoyningen-Huene, P. (1990): Kuhn's conception of incommensurability. *Studies in History and Philosophy of Science* **21**, 481–492.

Jung, C.G. (1969): On the nature of the psyche. In *Collected Works, Vol. 8*, second edition, Princeton University Press, Princeton.

Kuhn, T. (1983): Commensurability, comparability, communicability. In: Asquith, P. and Nickles, T. (eds), *Proceedings of the 1982 Biennial Meeting of the Philosophy of Science Association*, Philosophy of Science Association, East Lansing, pp. 669–688.

Pauli, W. (1950): Die philosophische Bedeutung der Idee der Komplementarität. *Experientia* **6**, 72–81, 1950. English translation in: Enz, C.P. and Meyenn, K. von (eds.), *Wolfgang Pauli. Writings on Physics and Philosophy*, Springer, Berlin 1994, pp. 35–42.

Pauli, W. (1952): Der Einfluss archetypischer Vorstellungen auf die Bildung naturwissenschaftlicher Theorien bei Kepler. In Jung, C.G. and Pauli, W. (eds.), *Naturerklärung und Psyche*, Rascher, Zürich, pp. 109–194. English translation in: Enz, C.P. and Meyenn, K. von (eds.), *Wolfgang Pauli. Writings on Physics and Philosophy*, Springer, Berlin 1994, pp. 219–279.

Primas, H. (1977): Theory reduction and non-Boolean theories. *Journal of Mathematical Biology* **4**, 281–301.

Primas, H. (2007): Non-Boolean descriptions for mind-matter problems. *Mind and Matter* **5**, 7–44.

Primas, H. (2008); Complementarity of mind and matter. *This volume.*

Quine, W.V. (1969): Ontological relativity. In *Ontological Relativity and Other Essays*, Columbia University Press, New York, pp. 26–68.

Römer, H. (2004): Weak quantum theory and the emergence of time. *Mind and Matter* **2**(2), 105–125.

Römer, H. (2006): Complementarity of process and substance. *Mind and Matter* **4**(1), 69–89.

Ruelle, D. (1989): The thermodynamic formalism for expanding maps. *Communications of Mathematical Physics* **125**, 239–262.

Seager, W. (2008): A new idea of reality: Pauli on the unity of mind and matter. *This volume.*

Shalizi, C.R. and Moore, C. (2003): What is a macrostate? Subjective observations and objective dynamics. Available at `xxx.lanl.gov/pdf/cond-mat/0303625`.

Slife, B. (2000): Are discourse communities incommensurable in a fragmented psychology? *Journal of Mind and Behavior* **21**, 261–271.

Sinai, Ya.G. (1968): Markov partitions and C-diffeomorphisms. *Functional Analysis and its Applications* **2**, 61–82.

Smolensky, P. (1988): On the proper treatment of connectionism. *Behavioral and Brain Sciences* **11**, 1–74.

Smolensky, P. (2006): Harmony in linguistic cognition. *Cognitive Science* **30**, 779–801.

Velmans, M. (2002): How could conscious experiences affect brains? *Journal of Consciousness Studies* **9**(11), 3–29.

Velmans, M. (2008): Psychophysical nature. *This volume.*

Viana, R.L. , Grebogi, G. , Pinto, S.E de S., Barbosa, J.R.R., and Grebogi, C. (2003): Pseudo-deterministic chaotic systems. *International Journal of Bifurcation and Chaos* **13**, 3235–3253.

Walach, H. and Römer, H. (2000): Complementarity is a useful concept for consciousness studies. A reminder. *Neuroendocrinology Letters* **21**, 221–232.

Westmoreland, M.D. and Schumacher, B.W. (1993): Non-Boolean derived logics for classical systems. *Physical Review A* **48**, 977–985.

Yanchar, S.C. and Slife, S.C. (1997): Pursuing unity in a fragmented psychology: Problems and prospects. *Review of General Psychology* **1**, 235–255.

Psychophysical Nature

Max Velmans

Department of Psychology, Goldsmiths College, University of London, UK,
`psa01mv@gold.ac.uk`

1 Introduction

In the present chapter we examine two quite distinct ways in which events that we normally think of as "physical" relate in an intimate way to events that we normally think of as "psychological". One intimate relation occurs in exteroception at the point where events in the world become events as-perceived. The other intimate relationship occurs at the interface of conscious experience with its neural correlates in the brain.

Normal exteroception involves an interaction between an event in the world (an event itself) and the perceptual/cognitive systems of an observer, which results in an event as-perceived. Such perceived events are the phenomena that form the basis of empirical science. Taken together, such perceived events also form our everyday "phenomenal worlds". Although we normally think of the world surrounding our bodies as the "physical world", science makes it abundantly clear that this perceived "physical world" is an appearance, whose nature is dependent not only on the nature of the world itself, but also on how information relating to that world is preconsciously processed by sense organs, perceptual systems and cognitive systems in the brain. The world that we actually see results from such preconscious observer-observed interactions, and can be very different in its apparent properties to the world as described by physics (in terms of quantum mechanics, relativity theory, and so on). Given this, is the world that we perceive "physical", "psychological" or somewhere in between ?

And this, in turn, raises a second question. Given the dependence of the perceived world on its proximal neural causes and correlates within the brain (as well as on events in the external world itself), what exactly is the ontology of this phenomenal world and its relationship to what is going on within the brain ? Is this perceived or experienced world nothing more than a brain state ? Is it something quite different to a brain state ? Or is it something in between ?

To answer these questions we have to grapple with one of the most fundamental issues for consciousness studies: How does consciousness relate to the brain and the physical world? I have dealt with many aspects of this and related issues in the “reflexive monism” that I develop in my book “Understanding Consciousness” (Velmans, 2000) and in various papers such as Velmans (1990, 2007, 2008). As is the case with consciousness studies in general, my own approach to these relationships has been largely guided by how consciousness, brain and the surrounding world manifest macroscopically, for example in the empirical findings of psychology, neuroscience, and classical physics.

However, with the recent availability of the unpublished writings of Wolfgang Pauli (Atmanspacher and Primas, 2006), it has become apparent that there are some interesting points of convergence, as well as some points of divergence, with some of Pauli’s prescient thoughts about the “psychophysical” nature of the microworld, that derive from his attempts to understand its nature via quantum mechanics. Most of these points of convergence and divergence have to do with the precise relationship of experienced (psychological) phenomena to their physical correlates in the brain, so this will be the main focus of the present chapter. However, normal exteroception is triggered by events in the world interacting with brain-based perceptual/cognitive systems that result in experienced phenomena which represent those triggering events in the world – and questions can also be asked about how the ontology of experienced phenomena relates to the events that they represent in the world. As this ontology provides a context for the later, more detailed discussion of how experienced phenomena relate to their neural correlates in the brain, I will briefly discuss this first.

2 Is the Perceived World “Physical”, “Psychological” or Somewhere in Between?

The ambiguous physical/psychological nature of perceived phenomena can best be understood in terms of the contrasts between three basic ways of making sense of how experiences and brains relate to the external physical world, known as dualism, materialist reductionism, and reflexive monism.

The classical view, which many of us intuitively adopt, is a form of dualism shown in Figure 1 below. This assumes perception to involve a simple, linear, causal sequence (viewed from the perspective of an external observer E). Light rays travelling from the physical object (the cat as-perceived by E) stimulate the subject’s eye, activating her optic nerve, occipital lobes, and associated regions of her brain. Neural conditions sufficient for consciousness are formed, and result in a conscious experience (of a cat) in the subject’s mind. This model of visual perception is, of course, highly oversimplified, but for now we are not interested in the details. We are interested only in where external physical objects, brains and experiences are placed.

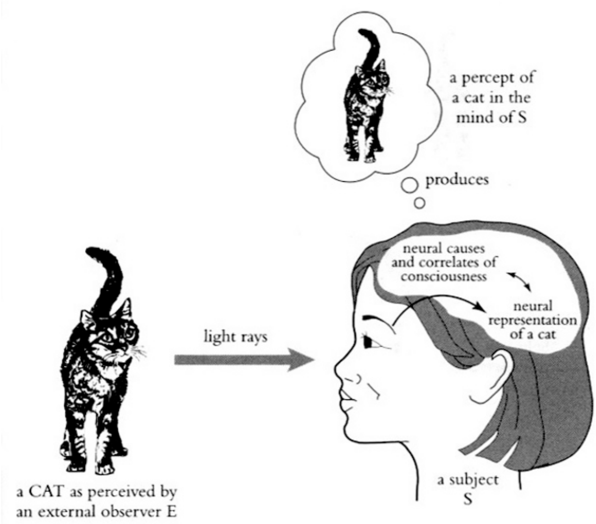


Fig. 1. A dualist model of perception

It will be clear that there are two fundamental "splits" in this model. Firstly, the contents of consciousness are clearly separated from the material world (the conscious, perceptual "stuff" in the upper part of the diagram is separated from the material brain and the physical cat in the lower part of the diagram). This conforms to Descartes' view that the stuff of consciousness (*res cogitans*, a substance that thinks) is very different to the stuff of which the material world is made (*res extensa*, a substance that has extension and location in space). Secondly, the perceiving subject is clearly separated from the perceived object (the subject and her experiences are on the right of the diagram and the perceived object is on the left of the diagram).

In short, on this dualist view, "physical phenomena" have an autonomous existence, location and extension out-there in space – and, although experiences of those phenomena (psychological phenomena) are influenced by physical events in the brain, they have a separate existence in the mind, which has neither location nor extension in space.

It will be apparent to those familiar with modern consciousness studies that a mind that has neither location nor extension in space does not fit easily into the unified, largely materialist explanatory system offered by modern science. As a consequence, 20th century Western philosophy and science commonly tried to "naturalize" dualism by arguing or attempting to show that conscious experiences are nothing more than states or functions of the brain. A reductionist model of visual perception is shown in Figure 2. The causal sequence in Figure 2 is the same as in Figure 1, with one added step. While reductionists generally accept that the subject's experience of a cat seems to be "in the mind", they argue that it is really a state or function of the brain. In

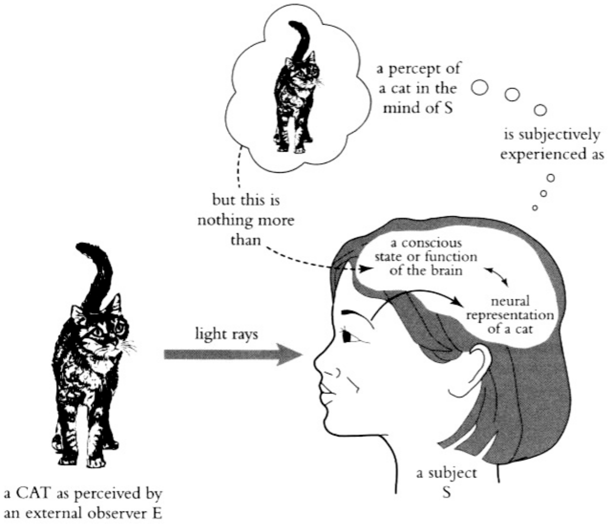


Fig. 2. A reductionist model of perception

short, the reductionist model in Figure 2 tries to resolve the conscious experience – physical world split by eliminating conscious experience or reducing it to something physical that E (the external observer) can in principle observe and measure. But reductionism retains the split (implicit in dualism) between the observer and the observed. The perceived object (on the left side of the diagram) remains quite separate from the conscious experience of the object (on the right side of the diagram).

On this reductionist view, "physical phenomena" have an autonomous existence, location and extension out-there in space, and experiences of those phenomena (psychological phenomena) are not just influenced by physical events in the brain, but *literally* are physical representations in the brain (of physical phenomena out-there in the world) that have their own distinct location and extension.

It will be apparent that dualism and reductionism present sharply conflicting views of the way that conscious (psychological) phenomena relate to physical phenomena – but, as the strengths and weaknesses of these positions have been extensively debated in the literature and as I have given extensive evaluations of both views of consciousness in Velmans (1998a; 2000, Chap. 2–5), I will not repeat this here. For the purposes of the present chapter, we simply need to note that both dualist and materialist explanations of conscious phenomenology claim its ontology to be very different to its appearances, which makes it difficult to explain the subtle ways in which conscious phenomenology appears to relate to events in the external world and in the brain. For example, contrary to dualism, nearly all experienced events appear to have both location and extension in space – yet contrary to materialist

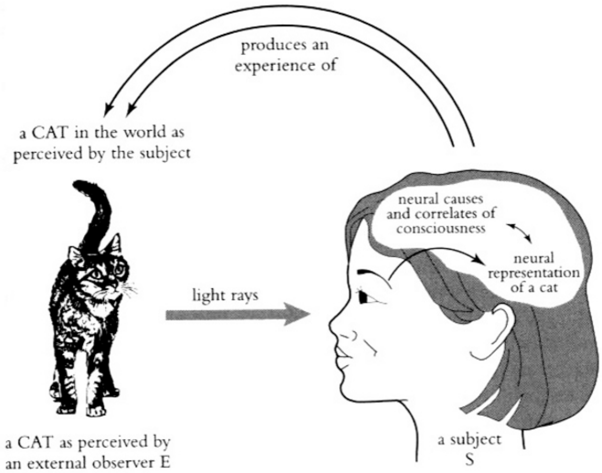


Fig. 3. A reflexive model of perception

reductionism, few experienced events appear to be located and extended in the brain. To give a few obvious cases, if one stubs one's toe one experiences pain, but the pain seems to be in the toe, not "nowhere" or "in the brain" – and if one looks at this print, it seems to be out here in space, but there does not seem to be some added "experience of print" in the mind or brain!

For the purposes of this chapter I will take it for granted that to deal with the subtleties of how conscious phenomenology relates to the brain and external world one has to start with an accurate description of that phenomenology. For this reason, I will focus on a reflexive monist view of consciousness in what follows. In what way does this offer a more accurate phenomenology ? The essential way in which it differs from both dualism and materialist reductionism is illustrated by the reflexive model of perception shown in Figure 3. In most respects Figure 3 is the same as Figures 1 and 2. As before, there is a cat in the world (perceived by E) that is the initiating stimulus for what S observes, and the neural causes and correlates of S's experiences are, as before, located in S's brain. The only difference relates to the way that the model represents S's experience. According to dualists, S's experience of a cat is "nowhere"; according to reductionists, S's experience of a cat is in her brain; according to the reflexive model, both of the former models misdescribe what S actually experiences. If you place a cat in front of S and ask her to describe what she experiences, she should tell you that she sees a cat in front of her in the world – and she has no additional experience of a cat "nowhere" or "in her brain."

It should be easy to grasp the essence of this. The objects that we experience seem to be out there in the world, not in our head or brain. But this immediately presents us with a problem. Given that the neural causes and correlates of what we experience are in the head or brain, how do the experiences get to be out there – an effect that I refer to as “perceptual projection”.

3 Perceptual Projection

As I have discussed the scientific status of perceptual projection elsewhere (Velmans, 1990; 2000, Chap. 6; 2008) I will give only a brief introduction here. Crucially, perceptual projection refers to an empirically observable effect, for example, to the fact that this print seems to be out here on this page and not in your brain. In short, perceptual projection is an effect that requires explanation; perceptual projection is not itself an explanation. We know that nonconscious processes within the brain produce consciously experienced events, which may be subjectively located and extended in the phenomenal space beyond the brain. We also know that this effect is subjective, psychological and viewable only from a first-person perspective. Nothing physical is projected from the brain.

While we do not have a full understanding of how perceptual projection occurs, there is a large experimental literature on the cues that are used to construct perception of distance and location. One example is the way three-dimensionality is gradually constructed by the brain from cues laid out in two dimensions in stereograms[1] or, more immediately, from perspective cues displayed on a two-dimensional surface in the way shown in Figure 4.

Virtual realities (VR) provide added ways of studying perceptual projection in operation. In virtual reality one appears to interact with a virtual world outside one’s body although there is no actual (corresponding) world there. So, in this situation, there is no danger of confusing the appearance of the virtual world with an actual world that one sees. Yet, objects in a VR world appear to have three-dimensional location and extension. Virtual objects can also be given what appear to be classical “physical” properties such as “hardness”; for example, the observer may wear a gauntlet on her hand which is programmed to resist closing around a visually perceived, virtual object, making the latter feel “solid”.

In truth, however, there is nothing solid there. Such virtual appearances do not fit easily into either a dualist or reductionist understanding of consciousness (see Velmans, 1998b). In spite of being nothing more than appearances, they do not appear to be either “nowhere” or “in the brain”. But they fit naturally into the reflexive model. When visual inputs from screens in VR headsets are appropriately co-ordinated with head and body movements, they provide information which resembles that arriving from actual objects in the world. The mind/brain models this information in the normal way, and constructs

---

[1] One can easily create stereograms of one’s own with the assistance of Kasuhiko Kondo’s program at `www.eyetricks.com/stereograms/onlinetools/stereocreator.htm`

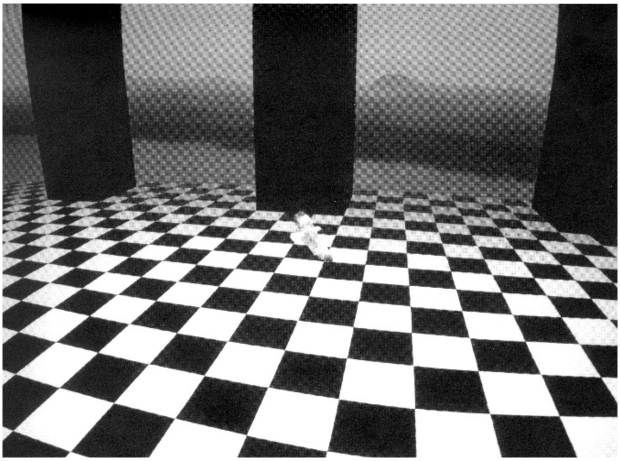


Fig. 4. A painting that uses radial perspective (developed and painted by Peter Cresswell). If one scans this picture through a rolled up tube (avoiding the edges), a strong perception of depth will result, even if the picture is inspected with only one eye.

what it normally constructs when it receives such input – a perceived, phenomenal world located and extended in the three-dimensional space beyond the body surface!

4 Consequences for the Perceived "Physical World"

What are the consequences of thinking about the perceived world in this reflexive way ? Although we normally think of the objects that we see around us as being "physical", they are in another sense "psychological". This is because they are the objects as they appear to us and not the objects as they are in themselves. Although it is natural (and, in a way, correct) to think of these appearances as the appearances of the objects themselves, the fact that they appear to us in the way that they do depends as much on the operation of our own perceptual systems as it does on the nature of the objects themselves. If we did not have color vision they would not appear colored in the way that they do, if we did not have tactile receptors they would not feel solid in the way that they do, and so on. Conversely, modern physics (quantum mechanics, relativity theory, etc.) offers descriptions of the deeper nature of these objects that are very different to their surface appearances.[2]

---

[2] It follows that once an object appears to us (once it has an appearance) the perceptual processing in our own mind/brain that contributes to that appearance has already operated. In short, the world as it appears to us (the phenomenal world) is the end product of our current (and very recent) perceptual processing

This convergence of psychological with physical phenomena is self-evident in situations where the same phenomenon can be thought of as either "physical" or "psychological", depending on one's interest in it. At first glance, for example, a visual illusion of the kind shown in Figure 5 might seem to present difficulties, for the reason that physical and psychological descriptions of this phenomenon conflict. Physically, the figure consists entirely of squares, separated by a horizontal line. But subjectively, the line seems to tilt down to the left, and the squares do not seem to be entirely square. However, these physical and psychological descriptions result from two different observation procedures. To obtain the physical description, an experimenter E can place a straight edge against each line, thereby obscuring the cues responsible for the illusion and providing a fixed reference against which the curvature and orientation of the line can be judged. To confirm that the line is actually straight, other experimenters (E_1 to E_n) can repeat this procedure. In so far as they each observe the line to be straight under these conditions, their observations are public, intersubjective and repeatable.

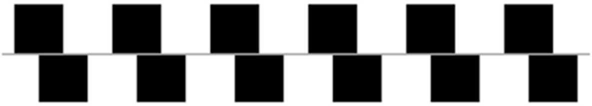


Fig. 5. In what way does the central line tilt ?

But, the fact that the line appears to be bent and to tilt to the left (once the straight edge is removed) is similarly public, intersubjective and repeatable (amongst subjects S_1 to S_n). Consequently, the illusion can be investigated using relatively conventional scientific procedures, in spite of the fact that the *illusion* is unambiguously *mental*. One can, for example, simply move the straight edge outside the figure making it seem parallel to the central line – thereby obtaining a measure of the angle of the illusion.

This basic relationship between what is physical and what is psychological applies not just to perceived lines but also to the entire, external, visually perceived world, and may be summarized in the following way: although we commonly assume the perceived three-dimensional external world to be "physical" and consequently something that is *separate from* consciousness, it

---

and not the cause of that processing. The true initiating cause of our perceptual processing in this situation is the object (or world) itself. Consequently, although in Figure 3 the initiating cause of what S perceives is labelled as "a cat as perceived by E" the true initiating cause of what S perceives (and of what E perceives) are light reflectances from the cat itself. It is labelled as "a cat as perceived by E" in Figure 3 for the reason that the figure represents the situation as viewed from E's perspective, and the cat itself appears as phenomenal cat when viewed by E (just as it does when viewed by S).

is actually *part of* what we experience, and therefore part of the contents of consciousness. This applies equally to those components of the phenomenal world that we normally think of as "physical phenomena" – and in this special sense the existence and nature of "physical phenomena" are dependent on the existence and nature of conscious experience.

It is important to note, however, that this conclusion, based on observed macrophenomena and classical physics, is tangential to the controversy in quantum mechanics about whether a measurement suffices to collapse superimposed quantum states into a single realized state, or whether human consciousness is somehow required. According to the reflexive model of perception, observed phenomena *represent* things themselves, but are not identical to them. Consequently, at macroscopic scales, things themselves can exist whether or not they are consciously perceived.[3] That said, once a phenomenon is observed, the form that the phenomenon will take is dependent not just on the nature of the observed but also on the nature of the observation arrangements, measuring equipment, and perceptual/cognitive processes available to the observer – and that applies equally within classical physics and quantum mechanics.

5 How the Perceived Physical World Relates to Information Processing in the Brain

It should be apparent from the above that, in cases of normal exteroception, questions about how conscious experience relates to its neural correlates, translate into questions about how an individual's *phenomenal world* relates to its neural correlates (for the simple reason that in terms of *phenomenology* an individual's "conscious experience" and their "phenomenal world" are one and the same). As noted above, the external phenomenal world, viewed from the perspective of an individual observer, appears to have a three-dimensional spatial extension and curvature with a definable topology that is different in a number of respects to that of measured Euclidian space – but that is nevertheless situated outside of the brain. By definition, however, the neural correlates of that experienced world must be located in some neural state space that is located inside the brain. In principle therefore it should be possible to specify the topological mapping of phenomenal space onto neural state space; see for example Lehar (2003) for an initial attempt.

Given that the search for the neural correlates of different conscious experiences is still very much a work in progress (see, for example, Rees and Frith, 2007; Crick and Koch, 2007), can anything general be said about them? By definition, correlates accompany or co-occur with given conscious experiences. This differentiates them from the antecedent causes of consciousness (such as the operation of selective attention, binding, etc.) which may be thought of

---

[3] See further discussion of idealism versus realism in Velmans (1990, 2000).

as the necessary and sufficient prior conditions for consciousness in the human brain. And, although we know little about the physical nature of these correlates, there are three plausible, functional constraints imposed by the phenomenology of consciousness itself.

1. *The representational constraint:* Normal human conscious experiences are representational (phenomenal consciousness is always *of* something). Given this, it is plausible to assume that the physical correlates of such experiences are representational states.
2. *The identical referent constraint:* A representational state must represent *something.* For a given physical state to be the correlate of a given experience it is plausible to assume that it represents the *same* thing.
3. *The information preservation constraint:* For a physical state to be the correlate of a given experience, it is reasonable to suppose that it has the same "grain". That is, for every discriminable attribute of experience there will be a distinct, correlated, physical state. As each experience and its physical correlate represent the same thing, it follows that each experience and its physical correlate encode the same information about that thing. That is, they are representations with the same *information structure.*

Although these assumptions have not always been made explicit in theories of consciousness they are largely taken for granted in psychological theory. Psychophysics, for example, takes it for granted that for any discriminable aspect of experience (a just noticeable change in brightness, color, pitch, and so on) there will be a correlated change in some state of the brain. The same is true for the more complex contents of consciousness, in the many cognitive theories that associate (or identify) such contents with information stored in primary (working) memory, or information at the focus of attention. The assumption that experiences and their physical correlates encode identical information also marks an important point of convergence between otherwise divergent theories about the nature of consciousness. This assumption is implicit, for example, in eliminativist, and reductionist theories of consciousness (such as Dennett, 1994, and Sloman, 1997). It is also explicit in the "naturalistic dualism" developed by Chalmers (1996) and in the dual-aspect theory developed in Velmans (1991a,b, 1996) which I elaborate below.

It is important to stress that having an identical referent and information structure does not entail *ontological identity* (as eliminativists and reductionists tend to assume). A filmed version of the play "Hamlet", recorded on videotape, for example, may have the same sequential information structure as the same play displayed in the form of successive, moving pictures on a TV screen. But it is obvious that the information on the videotape is not ontologically identical to the information displayed on the screen. In this instance, the same information is embodied in two different forms (patterns of magnetic variation on tape versus patterns of brightness and hue in individual pixels on screen) and it is displayed or "formatted" in two different ways (only the latter display is in visible form). Consequently the choice between eliminativism, reductionism, dualism, and dual-aspect theory has to be made on some other grounds, for example on the basis of which theory accounts for *all* the observable evidence in the most elegant way.

6 Creeping up on Consciousness

Eliminativism and reductionism assume that once one has identified the physical causes and correlates of consciousness in the brain, viewed from a third-person perspective, there is nothing else to understand or explain. For them, the neural correlates of consciousness (or the information structure they embody) are consciousness itself. However, this view is *inconsistent* with our first-person evidence about what experiences are like. Consequently its protagonists attempt to denigrate the utility, reliability or even the reality of first-person experience. Given the apparent importance of first-person experience to everyday human life, many find such manoeuvres evasions rather than explanations.

However, if one does not deny the reality of first-person experience, one is left with a conceptual problem. Once one arrives at the end of a third-person physical or functional account of how a brain or other system works one still needs some credible way to cross the “explanatory gap” to conscious experience. Luckily, in the human case, this is not really a *practical* problem, for the reason that we naturally have access to *what lies on both sides of the gap*. We can observe what is going on in the brains of others or in our own brain from an external third-person perspective (via exteroception, aided by a little physical equipment). And we naturally have first-person access to what it is like to have the experiences that accompany such observable brain activity. For many explanatory purposes we just need to switch from one perspective to the other at the appropriate place, and add the first-person to the third-person story in an appropriate way. In psychophysics, for example, one can examine the neural causes and correlates of a given experience in the brain viewed from a third-person perspective. But to complete the causal story, one then has to switch to the subject’s first-person perspective to get an account of the perceptual effect.

Note that this common-sense account of how the “explanatory gap” is crossed in practice is nonreductive. Third-person evidence about the workings of the brain retains its full privileged status (about the workings of the brain), and first-person evidence about what it is like to have a given experience retains its full privileged status (about the nature of experience). That said, neither third- nor first-person accounts are incorrigible. Once observations or experiences made from either perspective are translated into *descriptions* (observation statements or phenomenological descriptions) there is always a measure of interpretation required. Interpretation and abstraction is also required to translate such observations/experiences into general *descriptive systems*, typologies, and “maps” – and further inference and interpretation is required to translate first- or third-person evidence into a *theory about* the workings of mind, consciousness or brain. In all this, the normal rules of scientific engagement apply.

7 The Relation Between First-Person Descriptions of Experience and Third-Person Descriptions of Their Physical Correlates

While *perspectival switching* from a third-person account of neural events to a first-person account of correlated experiences allows one to cross the "explanatory gap" we still need to understand how such accounts relate to each other. Suppose, for example, I ask you to look at a cat out in the world while I examine the physical correlates of what you see in your brain (in the way shown in Figure 3). While I examine your brain I simply report what I see (whether or not I am aided by sophisticated equipment), and while you are looking at the cat you simply report what you see. In this situation, we both experience something out in the world that we would describe as "physical". You have a visual experience of a cat, located beyond your body, out in the world. I have a visual experience of the physical correlates of the cat that you see, beyond my body, in your brain.

Following the representational, identical referent, and information preservation constraints suggested above, what you and I see relates to each other in a very precise way. What you see is a phenomenal cat – a visual representation containing information about the shape, size, location, color and texture of an entity that currently exists out in the world beyond your body surface. What I see is the same information (about the cat) encoded in the physical correlates of what you experience in your brain. That is, the information structure of what you and I observe is identical, but it is displayed or "formatted" in very different ways. From your point of view, the only information you have (about the entity in the world) is the phenomenal cat you experience. From my point of view, the only information you have (about the entity in the world) is the information I can see encoded in your brain. The way your information (about the entity in the world) is displayed appears to be very different to you and me for the reason that the "observational arrangements" by which we access that information are entirely different. From my external, third-person perspective I can only access the information encoded in your neural correlates by means of my visual or other exteroceptive systems, aided by appropriate equipment. Because you *embody* the information encoded in your neural correlates and it is already at the interface of your consciousness and brain, it displays "naturally"[4] in the form of the cat that you experience.

---

[4] I assume that it is simply a "natural" empirical fact about the world that certain physical events in the brain (the correlates of consciousness) are accompanied by

You experience a cat, rather than your neural encodings of the cat, for the reason that it is the information *about the world* (encoded in your neural correlates) that is manifest in your experience rather than the embodying format or the physical attributes of the neural states themselves. As with the TV analogy above, the information encoded on videotape is displayed in the form of a picture on a screen without the magnetic fluctuations on the videotape or the tape itself being displayed upon the screen. I observe/experience the neural encodings of the cat in your brain (rather than the cat) for the simple reason that my visual attention is focused on your brain, not the cat. If I wanted to experience what you experience, I would have to shift my attention (and gaze) away from your brain to the cat.

From my "external observer's perspective", can I assume that what you experience is really nothing more than the physical correlates that I can observe? From my external perspective, do I know what is going on in your mind/brain/consciousness better than you do? Not really. I know something about your mental states that you do not know (their physical embodiment). But you know something about them that I do not know (their manifestation in experience). Such first- and third-person information is *complementary.* We need your first-person story and my third-person story for a complete account of what is going on.[5]

If I cannot reduce your story about what you experience to my story about its neural correlates (or *vice versa*) without loss, are we forced into the conclusion that experiences and their neural correlates are fundamentally different entities or substances? No. While dualism accepts the reality of first-person experience, it misdescribes its phenomenology. Descartes likens *all* experiences to "thoughts" (*res cogitans*) which, if they are verbal thoughts, take the form of "inner speech". However, most of what we experience has little resemblance to thoughts. For example, the way our bodies look and feel is quite unlike the phonemic imagery of inner speech, and the same is true of the look, sound, touch, taste and smell of entities in the external world such as phenomenal cats. Nor does splitting the universe into two, incommensurable (material and mental) substances help us to understand the *intimate relationship* of consciousness to matter.

The above analysis rather suggests a seamless universe, of which we are an integral part, which can be known in two fundamentally different ways. At the interface of consciousness and brain it can be known in terms of how it appears (from the outside) and it terms of what it is like to be that universe (from the inside). This is *ontological monism* combined with *epistemological dualism.*

---

experiences. In short, this relationship follows some natural law, however mysterious this presently seems.

[5] An introduction to "psychological complementarity" is given in Velmans (1991a), Sect. 9.3; Velmans (1991b), Secs. 8 and 9; Velmans (1993, 1996).

8 The Nature of Mind

What dwells within the "explanatory gap"? Ontological monism combined with epistemological dualism assumes that there must be some thing, event or process that one can know in two complementary ways. There must be something that grounds and connects the two views we have of it. Let us call this the "nature of mind".

If mind grounds and unifies the first- and third-person views we have of it, what can we conjecture about its nature?

1. Insofar as conscious experiences are of something or about something, it is reasonable to suppose that they, and their neural correlates, encode information. If so, the mind encodes information.
2. To the extent that brain activities and accompanying experiences are fluid and dynamic, the mind can be described as a process, developing over time.

Taken together, these points suggest that mind can be thought of as a form of information processing – and the information displayed in experiences and their physical correlates can be thought of as two manifestations of this information processing. However, this does not fully specify the ontology of the mind. Information processing needs to be encoded in some medium that is capable of carrying out that processing. Given this, what kind of "medium" is the mind?

One can give a very short list of the observable facts:

1. In the human case, minds viewed from the outside seem to take the form of brains (or some physical aspect of brains).
2. Viewed from the perspective of those who embody them, minds take the form of conscious experiences.

If first- and third-person perspectives (on the mind) are complementary and mutually irreducible, then the nature of the mind is revealed as much by how it appears from one perspective as the other. If so, the nature of mind is not *either* physical *or* conscious experience, it is at once physical *and* conscious experience. For lack of a better term we may describe this nature as *psychophysical*. If we combine this with the features above, we can say that mind is a psychophysical process that encodes information, developing over time.

At present, there is little more about "what dwells within the explanatory gap" that can be said with confidence. However, there are some useful pointers to what a more complete theory of mind would look like, that can be drawn from other areas of science. At the macrocosmic level, the relation of electricity to magnetism provides a clear parallel to the form of dual-aspect theory suggested above. If one moves a wire through a magnetic field, this produces an electrical current in the wire. Conversely, if one passes an electrical current through a wire, this produces a surrounding magnetic field. But it does not make sense to suggest that the current in the wire is nothing more than the surrounding magnetic field, or *vice versa* (reductionism). Nor is it accurate to suggest that electricity and magnetism are energies of entirely different kinds that happen to interact (dualist interactionism). Rather these are two manifestations (or "dual-aspects") of *electromagnetism*, a more fundamental energy that grounds and unifies both, described with elegance by Maxwell's laws.

The struggle to find a model or even a form of words that somehow captures the dual-aspect nature of mind is also reminiscent of wave-particle complementarity in quantum mechanics – although this analogy is far from exact. Light either appears to behave as electromagnetic waves or as photon particles depending on the "observation arrangements". And it does not make sense to claim that electromagnetic waves really *are* particles (or *vice versa*). A complete understanding of light requires both complementary descriptions – with consequent struggles to find an appropriate way of characterizing the nature of light which encompasses both descriptions ("wave-packets", "electron clouds", and so on). This has not prevented physics from developing very precise accounts of light viewed either as waves or as particles, together with precise formulae for relating wave-like properties (such as electromagnetic frequency) to particle-like ones (such as photon energy). If first- and third-person accounts of consciousness and its physical correlates are complementary and mutually irreducible, an analogous *psychological complementarity principle* might be required to understand the nature of mind.

9 Similarities and Differences to Pauli

This dual-aspect theory of information developed from entirely psychological considerations in Velmans (1991a,b, 1993, 1996, 2000) has some interesting similarities and differences to one later developed in the philosophy of David Chalmers (1995, 1996).[6] However, its close relationship to Pauli's thoughts on the subject, written in previously unpublished letters, is even more striking.

---

[6] As I have reviewed these similarities and differences in Velmans (1995, 1998c, 2000) I will not enter into a discussion of them here. Briefly, Chalmers and I agree that: (1) Phenomenal experiences and their neural or functionally defined correlates share the same information structure. (2) Phenomenal experiences are not reducible to their neural or functionally defined causes and correlates. (3) It should be possible to relate conscious experiences to their correlates via bridging laws. (4) Consciousness is a basic property of the universe. Our theories differ in that: (1) Although Chalmers sometimes calls his analysis "double-aspect theory" (his own term for "dual-aspect" theory), he usually, more accurately, calls it "naturalistic dualism". (2) The reason for the latter term being more accurate is that in Chalmers' theory, there is nothing deeper (ontologically unifying) such as "the nature of mind" of which experiences and their neural correlates are aspects. Consequently, (3) first- and third-person accounts are not "complementary" accounts of a psychophysical mind. Rather, (4) according to Chalmers, experiences "supervene" on their physical correlates – which conflicts with his contention that

In particular, consequent on his discussions with C.G. Jung, Pauli posited a similar, underlying psychophysical reality of which mind and physical matter are complementary aspects. As he wrote (Pauli, 1952; cited by Atmanspacher and Primas, 2006):

> "For the invisible reality of which we have small pieces of evidence in both quantum physics and the psychology of the unconscious, a symbolic psychophysical unitary language must ultimately be adequate, and this is the far goal which I actually aspire. I am quite confident that the final objective is the same, independent of whether one starts from the psyche (ideas) of from physis (matter). Therefore, I consider the old distinction between materialism and idealism as obsolete."

And, on psychological complementarity, Pauli (1952) wrote : "It would be most satisfactory if physis and psyche could be conceived as complementary aspects of the same reality." In their commentary on Pauli, Atmanspacher and Primas (2006, p. 28) make it clear that this amounts to ontological monism (a *unus mundus*) combined with epistemological dualism:[7]

> "The concept of the *unus mundus* provides and ontological level of description without any split of mental and material domains, which is more fundamental than the descriptive level with split domains. One can address the transition from the fundamental level to that with mind and matter separated in terms of emergence, if one thinks of it as an emergence of the distinction of mind and matter (rather than the emergence of mind from matter)."

Given our very different points of departure (quantum mechanics and Jungian depth psychology versus the psychology of perception) there are of course differences in emphasis between Pauli and Atmanspacher and Primas on the one hand, and my own analysis, briefly introduced above. Pauli and Atmanspacher and Primas, for example, give some thought to the formative principles, Platonic universals or archetypes that might underly both psychic and material manifestations of the *unus mundus*, and Atmanspacher and Primas (2006) suggest that mathematical formalism governing symmetry and symmetry breaking might provide a useful entry to an understanding of the way that formlessness might give rise to form.

---

consciousness is "basic" (if experiences supervene on the physical, in what sense are they "basic" ?). Finally (5) in Chalmers' formulation, there is no account of psychophysical causation, a major concern of my own dual-aspect theory of mind (see Velmans, 1993, 1996, 2000, 2002a,b).

[7] Note that Pauli, Atmanspacher and Primas, and Velmans agree on these points, but differ from Chalmers. In Chalmers' theory there is neither a "psychophysical *unus mundus*", nor a "complementarity" between mind and physical matter – and, rather than the "emergence of a distinction between mind and matter" from an underlying unity, Chalmers is committed to the view that conscious experiences "supervene" on physical states (see note 6 above).

My own concern with the way that conscious experiences relate both to their neural correlates and to the external entities and events that they represent has led to a focus on *information* rather than principles of formation. However there is no conflict between these differences in emphasis. Information needs to be formatted or encoded in some medium, so an understanding of how information emerges from the *unus mundus* has to be combined with an understanding of the emergence of form.

That said, there are also some genuine differences between reflexive monism and Pauli's thought. For example, there is no hint of the "reflexive" aspect of reflexive monism in Pauli's writings. Consequently, in the following extract he expresses the belief that only modern physics (in the form of quantum mechanics) offers an avenue for unifying the psychological with the physical (Pauli, 1953):

> "It is true that the distinction of 'physical' and 'psychic' is inevitable in the empirical world of phenomena, and it was the mistake of the alchemists to apply a monistic (neutral) language to concrete chemical processes. But since matter has now turned into an abstract, invisible reality for the modern physicist, the prospects for a psychophysical monism have become much more auspicious."

In contrast, reflexive monism suggests that perceived phenomena themselves can be both psychological (insofar as they are appearances) and physical (in so far as they are appearances of independently existing things) – although it is agreed that physical *appearances* have to be distinguished from the abstract realities described by modern physics, e.g. in the mathematical formalisms of quantum mechanics.

10 Similarities and Differences Between Physical and Psychological Complementarity

As noted above there are some genuine similarities between psychological complementarity (a way of understanding the relationship of conscious experiences to their neural correlates in the brain), and complementarity in quantum mechanics. In particular (1) Complementary observations are obtained from different observational arrangements; (2) Complementary descriptions are mutually irreducible; (3) For any complementary pair of observations of a given entity or event one needs descriptions of both observations for a complete account of the observable properties of that entity or event.

However, these tempting similarities should not obscure some genuine differences. In particular: (1) Complementary descriptions in physics are based on third-person observations, but complementary descriptions of phenomenal experiences and their neural correlates are normally based on, respectively, first-person and third-person observations; (2) Complementarity in physics is exclusive in the sense that making one observation of a complementary pair excludes the possibility of making the other, paired observation. However, complementarity in psychology is normally non-exclusive for the reason that a subject can have a given experience and report on it, while an external observer can simultaneously observe and report on the neural correlates of that experience.

Note however that, unlike physical complementarity, non-exclusive psychological complementarity relies on the possibility of simultaneous observations made by two independent observers (an external observer and a perceiving subject). A closer analogy with quantum mechanics may therefore be a hypothetical "autocerebroscope" experiment, in which an individual observer attempts to observe the neural correlates of his/her own *current* experience. In this situation, the neural correlates of the observer's visual experience are displayed in real-time in a visible form, for example on a monitor screen, and the observer simply looks at the screen. Note that, in principle, there should be no impediment to observing a visual on-screen representation of the neural correlates of one's own *past* visual experience (this is already possible, in limited ways, with imaging equipment). It may also be possible to shorten the delay between current experience and observations of its correlates within limits set by the processing and display time of the measuring system and the processing time of the visual system itself.

But, in real-time, it may be impossible in principle to observe the neural correlates of one's own *current* experience. Even if the delays in the system could be reduced to near zero, like a dog chasing its own tail, one would never quite catch up. In these circumstances psychological complementarity would be, *in this special sense*, exclusive. However, this still falls short of the exclusivity found in quantum mechanics. In quantum mechanics, measurement of one member of a complementary pair confines the accuracy of a *subsequent* (as well as simultaneous) measurement of the other member of that pair, but in psychological complementarity there would seem to be nothing to confine subsequent observation and measurement of the neural correlates of one's own current experience.

11 Conclusions

There appear to be interesting similarities between aspects of Pauli's thought, elaborated by Atmanspacher and Primas (2006), and reflexive monism about the best way to understand the mind-matter relationship at the interface of conscious experiences with their neural correlates in the brain. Given that reflexive monism is largely concerned with psychological issues, and that Pauli's concern is primarily with modern physics, it is significant that both arrive at the view that the relationship of conscious experiences to their neural correlates can be understood in terms of dual aspects of an underlying, unifying wholeness (or *unus mundus*) whose nature can best be described as "psychophysical". There also appear to be points of similarity between the "psychological complementarity" developed in reflexive monism and complementarity in physics.

While such similarities should not obscure other, genuine differences, for example in the "reflexive" aspect of reflexive monism, and in the non-exclusive nature of psychological complementarity, there appears to be a prospect of some genuine convergence on these fundamental issues between psychology and physics.

References

Atmanspacher, H. and Primas, H. (2006): Pauli's ideas on mind and matter in the context of contemporary science. *Journal of Consciousness Studies* **13**(3), 5–50.

Chalmers, D. (1995): Facing up to the problem of consciousness. *Journal of Consciousness Studies* **2**(3), 200–219.

Chalmers, D.J. (1996): *The Conscious Mind: In Search of a Fundamental Theory.* Oxford University Press, New York.

Crick, F. and Koch, C. (2007): A neurobiological framework for consciousness. In: Velmans, M., and Schneider, S. (eds.), *The Blackwell Companion to Consciousness.* Blackwell, New York, pp. 567–579.

Dennett, D.C. (1994): Instead of qualia. In: Revonsuo, A., and Kampinnen, M. (eds.), *Consciousness in Philosophy and Cognitive Neuroscience.* Lawrence Erlbaum, Hillsdale, NJ.

Lehar, S. (2003): Gestalt isomorphism and the primacy of subjective conscious experience. *Behavioral and Brain Sciences* **26**, 375–444.

Pauli, W. (1952): Der Einfluss archetypishcher Vorstellungen auf die Bildung naturwissenschaftlicher Theorien bei Kepler. In: Jung, C.G., and Pauli, W. (eds.), *Naturerklärung und Psyche.* Rascher, Zürich, pp. 109–194.

Pauli, W. (1953): Discussion remark to the talk "Isotopic spin and mass quantization" by A. Pais. *Physica* **19**, 887.

Rees, G. and Frith, C. (2007): Methodologies for identifying the neural correlates of consciousness. In: Velmans, M., and Schneider, S. (eds.), *The Blackwell Companion to Consciousness.* Blackwell, New York, pp. 553–566.

Sloman, A. (1997): Design spaces, niche spaces and the "hard" problem. Available online at `www.cs.bham.ac.uk/~axs/my-doings.html#designspace`

Velmans, M. (1990): Consciousness, brain, and the physical world. *Philosophical Psychology* **3**, 77–99.

Velmans, M. (1991a): Is human information processing conscious? *Behavioral and Brain Sciences* **14**, 651–669.

Velmans, M. (1991b): Consciousness from a first-person perspective. *Behavioral and Brain Sciences* **14**, 702–726.

Velmans, M. (1993): Consciousness, causality and complementarity. *Behavioral and Brain Sciences* **16**, 404–416.

Velmans, M. (1995): The relation of consciousness to the material world. *Journal of Consciousness Studies* **2**(3), 255–265.

Velmans, M. (1996): Consciousness and the "causal paradox". *Behavioral and Brain Sciences* **19**, 537–542.

Velmans, M. (1998a): Goodbye to reductionism. In: Hameroff, S., Kaszniak, A., and Scott, A. (eds.), *Towards a Science of Consciousness II: The Second Tucson Discussions and Debates.* MIT Press, Cambridge MA, pp. 45–52.

Velmans, M. (1998b): Physical, psychological and virtual realities. In: Wood, J. (ed.), *Virtual and Embodied Realities.* Routledge, London, pp. 45–60.

Velmans, M. (1998c): Review of D. Chalmers "The Conscious Mind". *Consciousness and Experiential Psychology* **1**, 14–17.

Velmans, M. (2007): Dualism, reductionism and reflexive monism. In: Velmans, M., and Schneider, S. (eds.), *The Blackwell Companion to Consciousness.* Blackwell, New York, pp. 346–358.

Velmans, M. (2008): Reflexive monism. *Journal of Consciousness Studies* **15**(2), 5–50.

Complementarity in Bistable Perception

Harald Atmanspacher[1], Thomas Filk[2], and Hartmann Römer[3]

[1] Institute for Frontier Areas of Psychology, Freiburg, Germany, `haa@igpp.de`
[2] Parmenides Center for the Study of Thinking, Munich, Germany, `Thomas.Filk@t-online.de`
[3] Institute of Physics, University of Freiburg, Germany, `hartmann.roemer@physik.uni-freiburg.de`

1 Introduction

The idea of complementarity already appears in William James' (1890a, p. 206) *Principles of Psychology* in the chapter on "the relations of minds to other things". Later, in 1927, Niels Bohr introduced complementarity as a fundamental concept in quantum mechanics. It refers to properties (observables) that a system cannot have simultaneously, and which cannot be simultaneously measured with arbitrarily high accuracy. Yet, in the context of classical physics they would both be needed for an exhaustive description of the system.

In contrast to the concept of a "complement" in mathematics, which refers to the negation of a proposition,[4] complementarity refers to properties that are not simply negations of each other. A nice example is mentioned by James (1890b, p. 284): "The true opposites of belief ... are doubt and inquiry, not disbelief." Disbelief would be the complement of belief in the Boolean sense, while doubt and inquiry are concepts that are complementary to belief. Another pertinent example for complementarity may be "learning" and "knowing" in data processing systems. In addition to James and Bohr, Wolfgang Pauli was one of those scientists who always thought that the idea of complementarity is significant far beyond the objectively measurable realms of physics.

In quantum mechanics, complementarity is mostly used in the context of observables such as "momentum" and "position" which are, technically speaking, non-commuting observables. Although complementarity soon became an important ingredient in the so-called Copenhagen interpretation of quantum theory, there exists no rigorous and unique mathematical definition of complementarity which all scientists agree upon. There are many definitions which all

---

[4] For instance, the complement of a set in Boolean set theory consist of those elements which are not elements of that set.

emphasize the non-commutativity of the corresponding mathematical objects but which differ with respect to more restrictive conditions.

Several years ago, Atmanspacher *et al.* (2002, 2006) formulated a mathematical framework for dealing with observations and measurements, which generalizes the framework used in quantum (and classical) mechanics so that applications in psychology and cognitive science become possible. The Necker-Zeno model for bistable perception, which will be explained in this contribution, marks a first success in this direction.

Note that it is not the aim of the generalized quantum theory to explain mental or cognitive phenomena in terms of quantum *physics*. The idea is rather to use elements of the *mathematical* framework of quantum theory (in particular those elements which appear when an observation of a system changes its state) and apply them to non-quantum (i.e. classical) physical systems, and eventually even to non-physical systems.

Concerning terminology, we will sometimes use the terms "classical" and "non-classical" in the following sense: The behavior of a system is called classical if an observation of the system has no (or, at least, negligible) influence on the state of the observed system. In this limit observations commute and we obtain a behavior that is typical for systems in classical physics. In those cases, however, for which an observation of a system has an unavoidable effect on the state of the observed system, we may encounter non-commutative observations and thus non-classical behavior. This is the domain for which the generalized quantum theory is intended. There may be different levels of non-classical behavior, from simple examples for non-commutative observables up to non-classical behavior manfesting itself in the violation of Bell's inequalities.

Bistable perception is a particularly suited scenario for applying the generalized formalism of quantum theory. In a first approximation, one has to distinguish only two different mental states corresponding to the two different representations of an ambiguous stimulus (such as, e.g., the Necker cube). Simple assumptions about the state dynamics between representations lead to the Necker-Zeno model proposed by Atmanspacher *et al.* (2004) and refined by Atmanspacher *et al.* (2008). This model not only accounts for the feature that switches of the representation cannot be avoided. It also predicts a quantitative relation between three different cognitive time scales: the time scale at which the sequence of perceived stimuli becomes undecidable, the time scale at which perceptions of stimuli become consciously accessible, and the time scale at which mental states in bistable perception switch.

This contribution reviews the basic ideas of the generalized quantum theory and its application to the bistable perception of ambiguous stimuli, the so-called Necker-Zeno model. In the following Sec. 2 we will describe the phenomenon of bistable perception and some relevant experimental data. Before we then introduce the Necker-Zeno model in Sec. 4, we will briefly sketch the main ideas of the generalized quantum theory according to Atmanspacher *et al.* (2002, 2006). Finally, in Sec. 5, we will speculate about a new idea to use so-called temporal Bell inequalities (see, e.g., Leggett and Garg, 1985; Mahler, 1994) as a test for non-classical behavior (in the sense indicated above) in mental systems.

2 Bistable Perception

The bistable perception of ambiguous stimuli such as the Necker cube is a well-known phenomenon in cognitive science (Kruse and Stadler, 1995; Long and Toppino, 2004). It refers to the effect that the mental state of subjects perceiving an ambiguous stimulus, e.g. an image which can be interpreted in two (or more) different ways, switches spontaneously between the two (or more) possible perceptions, often perspectivally different. The time between two successive shifts, i.e. the inverse *reversal rate*, will be called *dwell time.* Key predictions of the Necker-Zeno model refer to the functional dependence of the dwell time on experimentally controllable parameters.

The perception of ambiguous stimuli shares many features with another scenario called *binocular rivalry*, where two different unambiguous stimuli are offered each to one eye of the observer (Blake and Logothetis 2001). However, there are also important differences, in particular with respect to the issue of voluntary control over the reversal rate (Meng and Tong, 2004). In this contribution we we will not address binocular rivalry and restrict ourselves to the perception of ambiguous stimuli.

A simple and often used example for an ambiguous stimulus leading to bistable perception is the Necker cube (Fig. 1, left), a projection of the edges of a three-dimensional cube onto a plane. There are two ways to give this drawing a three-dimensional interpretation: either the front side is lower left or it is upper right (Fig. 1, right).

In experimental studies of Necker-cube perception, subjects are asked to direct their view onto a fixation cross in the center of the image and report, e.g. by pressing a button, whenever they perceive a "switch" of the perspective. Fig. 2 shows a typical switching behavior between states 1 and 2 as a function of time with dwell times in the range between 1 and 5 seconds. Typical dwell time distributions (over many trials) are similar to gamma distributions (Brascamp *et al.*, 2005; Atmanspacher *et al.*, 2008) with a mean dwell time T.

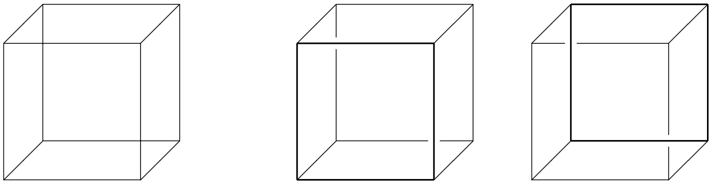


Fig. 1. The Necker cube (left) and the two ways how it can be interpreted (right).

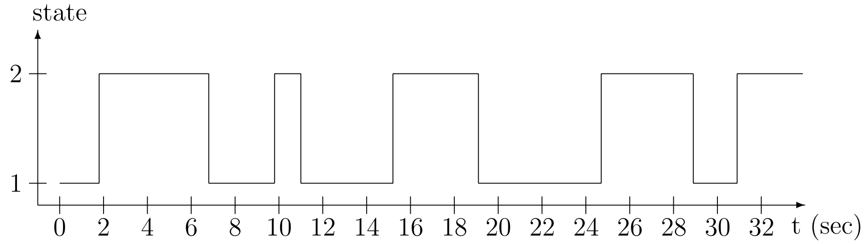


Fig. 2. Schematic representation of the bistable switching between states 1 and 2 as a function of time t.

It should be noted that inter-individual variations of dwell times can exceed intra-individual variations, sometimes by far. Different from usually found values of $T \approx 3$ seconds, Carter *et al.* (2005) reported that particular types of meditation can lead to dwell times that are increased up to several hundreds or even thousands of seconds.

The Necker-Zeno model explains why the mental state of subjects cannot be kept in one of the two representations of the Necker cube for an arbitrarily long time (see Sec. 4.2). On this basis, it provides a relation between the mean dwell time $T = \langle t \rangle$ and other cognitive time scales. A recently refined version of the model also gives correct predictions of the shape of the dwell time distribution and the cumulative dwell time probability (see Sec. 4.3).

3 Complementarity in Generalized Quantum Theory

The Necker-Zeno model was developed in the context of a generalized quantum theory (Atmanspacher *et al.*, 2002, 2006) and its application in cognitive science. In this section, we give a brief summary of the framework of generalized quantum theory with particular emphasis on possible formalizations of the concept of complementarity.

The developement of quantum theory in the 1920s and 1930s made it obvious that the assumption of a non-intervening or non-invasive measurement is unsuitable for systems with only a few elementary degrees of freedom. The concept of "observation" has to include the experimental fact that any observation may have an intrinsic and unavoidable influence on the state of the observed system and its associated observables. While in classical physics observables are mathematically represented as *functions* on the space of states (the phase space of a system), in quantum physics observables are represented as *operators* acting on the space of states.

Despite the significant differences between classical theory and quantum theory it turned out that both theories fit into one general algebraic framework – observables form a C*-algebra. In this framework the key distinction between classical and quantum physics is the distinction between the commutativity (classical physics) and the non-commutativity (quantum physics) of the observables. In both cases, a state is a positive, normalized, linear functional on the algebra of observables, associating to each observable its expectation value in that state.

Instead of referring to the algebra of all observables, one sometimes uses the structure of the set of propositions – observables with only 0 or 1 as a possible outcome of a measurement. Classical propositions form what is called a distributive lattice (corresponding to commutative observables), while quantum propositions form a non-distributive lattice (corresponding to non-commutative observables).

Even though the algebraic framework is general enough to comprise both classical and quantum physics, it contains some quite restrictive postulates. For instance, it is assumed that for any two observables A and B, also the sum $A + B$ is defined to be an observable, even though there exists no operational rule to derive the experimental protocol for the measurement of $A + B$ from the protocols for measuring A and B separately.

The generalized quantum theory provides a scheme for a mathematical representation of observables which is applicable to any system "which has enough internal structure to be a possible object of a meaningful study" (Atmanspacher *et al.*, 2002) In this respect, systems of interest in psychology and cognitive science are a particular challenge.

The complete and detailed axiomatic set-up of the generalized quantum theory has been published elsewhere (Atmanspacher *et al.*, 2002, 2006). Here it is sufficient to sketch the main ingredients. The basic elements of the theory are a set of states $\{z\}$ and a set of observables $\{A\}$. Observables act on the set of states as mappings, i.e., they can change the states. The main axioms are:[5]

- Observables have a *spectrum*, which is the set of all possible results of a measurement or an observation. The nature of the possible results of measurements remains unspecified. In particular, it is not required that the results can be expressed in terms of real numbers, or that results can be added or multiplied.
- Observables can be multiplied, which is related to the fact that their measurement can be performed in sequential order, i.e. in temporal succession. It should be noted, however, that not even in quantum mechanics the temporal succession of two observations is represented by the product of the corresponding observables (expressed by the operator or matrix product of linear mappings). The relation is more subtle: In quantum theory we assume that for each measurement of an observable A with duration t_1 there exists a time evolution operator $U_A(t)$ which describes the time evolution of the system (including the measuring device) during the measurement

---

[5] Other axioms, like the existence of an identity observable, a zero observable, and the existence of a zero state, are relevant for the developement of the mathematical structure but unimportant for a conceptual discussion.

process. The later measurement of a second observable B with duration t_2 is represented by a different time evolution operator $U_B(t)$. The time evolution corresponding to the temporal succession of both processes is represented by $U_B(t_2)U_A(t_1)$. These two operators do not commute if the corresponding operators A and B do not commute. The multiplication of observables in generalized quantum theory refers to the operators $U_A(t)$, $U_B(t)$ rather than to the operators A, B representing the observables in quantum mechanics.

In this framework, a number of significant features of quantum mechanics are missing: There is not necessarily a Hilbert space of states, there is no *a priori* probability interpretation, there is no unitary Schrödinger evolution describing the time evolution of states, etc. However, despite the small set of axioms, the conditions required for the observables yield several options to define complementarity:

1. A most general definition of complementarity refers to the commutativity of observables: two observables A and B are said to be complementary if the corresponding mathematical representations of these observables do not commute, i.e., if the results of temporally successive measurements of these observables depend on their temporal order.
   This definition of complementarity is quite weak: It may happen that two observables do not commute on a few exceptional states but commute on the majority of states.
2. More restrictively, we may call two observables A and B complementary if they do not commute on any state z which is not the zero state:

   $$ABz \neq BAz \quad \text{for all} \quad z \neq 0 \quad .$$

   In the framework of conventional quantum mechanics this implies that there are no states for which the two observables A and B assume definite values simultaneously. This corrolary is not necessarily equivalent with the definition according to generalized quantum theory and may, therefore, be considered as an alternative definition of complementarity.
3. A more restricted version of complementarity would require (apart from non-commutativity) that the observables A and B generate (by multiplication and any other additional operation which may be defined in special cases) the complete set of observables.
   For instance, this definition is satisfied if we think of position Q and momentum P for systems with only one degree of freedom in conventional quantum mechanics. It is not satisfied if there are several degrees of freedom, related to several particles and/or several dimensions and/or internal degrees of freedom. One can also give meaning to this definition of complementarity by generalizing the concept of complementarity from two observables A and B to two sets of observables $\{A_i\}$ and $\{B_i\}$, and requiring that both $\{A_i\}$ and $\{B_i\}$ commute among themselves and that A_i and B_i do not commute pairwise.

4. Finally, there are even more restrictive definitions of complementarity in quantum mechanics. The strongest definition of complementarity requires that the dispersion-free states[6] related to two observables A and B have a "maximal distance". While dispersion-free states can be defined in the context of generalized quantum theory as well, the concept of a "distance of states" needs more structure than provided by the framework of generalized quantum theory.

4 The Necker-Zeno Model

The Necker-Zeno model for bistable perception was first proposed by Atmanspacher *et al.* (2004). It is based on the same idea as the quantum Zeno effect introduced by Misra and Sudarshan (1977). Therefore we shall first describe the quantum Zeno effect in the form used for quantum systems proper. This does not mean that we want to hold a genuine quantum effect responsible for bistable perception. However, it will be shown that parts of the mathematical framework used for describing the quantum Zeno effect coincide (in the sense of generalized quantum theory) with the mathematical framework that is applicable to describe bistable perception.

4.1 The Quantum Zeno Effect

In a simple two-state model, the quantum Zeno effect can be described by the following ingredients:

1. Observations are represented by the operator

$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad .$$

   Immediately after an observation the system will be in one of the two eigenstates

$$\psi_1 = |+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{or} \qquad \psi_2 = |-\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad .$$

2. Without loss of generality, we may assume that the time evolution of the unperturbed system is generated by a Hamilton operator

$$H = g\sigma_1 = g \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad ,$$

   where g is some coupling constant related to the velocity at which the states change. The corresponding unitary operator of time evolution is given by

---

[6] A state z is called dispersion-free with respect to an observable A if the possible results of measurements of A in systems prepared in state z are precisely identical.

$$U(t) \;=\; e^{iHt} \;=\; \begin{pmatrix} \cos(gt) & i\sin(gt) \\ i\sin(gt) & \cos(gt) \end{pmatrix} \quad .$$

Now, the unperturbed time evolution of the system (without observation) can be compared with the case where repeated observations are performed at time intervals ΔT. The probability that the unperturbed system in, say, state $|+\rangle$ at time $t = 0$ is still found in state $|+\rangle$ at time t is given by:

$$w(t) \;=\; |\langle +|U(t)|+\rangle|^2 \;=\; \cos^2(gt) \quad . \tag{1}$$

The time scale $t_0 = 1/g$ characterizes the "relaxation" or "decay" of the unperturbed system into an observation eigenstate.

Considering repeated observations after time intervals ΔT, the joint probability that the system is in state $|+\rangle$ at $t = 0$ and at all subsequent observations until time $t = N \cdot \Delta T$ is given by:

$$w_{\Delta T}(t) \;=\; [\cos(g\,\Delta T)]^{2N} \;=\; \exp\left(2N \ln[\cos(g\,\Delta T)\,]\right) \quad . \tag{2}$$

For $g\,\Delta T \ll 1$ or $\Delta T \ll t_0$ we may expand the cosine and the logarithm and obtain:

$$w_{\Delta T}(t) \;\approx\; \exp\left(-g^2 \Delta T^2 \cdot N\right) \;=\; \exp -(t/T) \quad . \tag{3}$$

The decay time t_0 in the unperturbed case is now replaced by a slower time scale T for the decay of the system:

$$T \;=\; (g^2 \Delta T)^{-1} \;=\; t_0^2/(\Delta T) \quad . \tag{4}$$

For $\Delta T \to 0$ we find $T \to \infty$, i.e., for continuous observations the mean time for a change of the system from state $|+\rangle$ to state $|-\rangle$ tends to infinity: The system becomes frozen in $|+\rangle$ under the influence of the observations. Figure 3 illustrates the three time scales of the quantum Zeno effect.

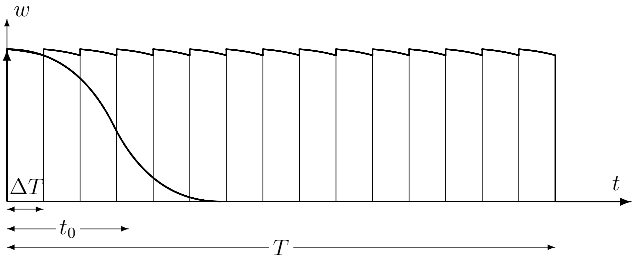


Fig. 3. Time scales of the quantum Zeno effect: ΔT is the time interval between successive observations of the system, t_0 is the time scale for the decay of the unperturbed (unobserved) system, and T is the mean decay time if the system is observed at time intervals ΔT.

4.2 The Necker-Zeno Model for Bistable Perception

The quantum Zeno effect can be related to the perception of ambiguous stimuli if the following correspondences are assumed:

1. The two states of the quantum Zeno effect correspond to the two possible representations of the ambiguous Necker cube (Fig. 1).
2. Without updates due to successive observations the mental representation "decays" with a probability which for small times t is given by

$$w(t) \;\approx\; 1 - g^2t^2 \,+\, \mathrm{O}(t^4) \quad . \tag{5}$$

   This expression coincides in lowest non-trivial order with the corresponding probability in the quantum Zeno effect (Eq. 1). The higher-order terms are not needed in the derivation of the quantum Zeno effect, and the oscillatory behavior for large t is not to be expected for the dwell time in bistable perception anyhow.
3. The observation of the stimulus provides "updates" at time intervals ΔT. After each update the mental state corresponds to one of the two possible three-dimensional representations of the Necker cube.

In the resulting Necker-Zeno model, two types of processes can be considered as complementary: (i) the bistable perception dynamics (formalized by σ_1) tends towards a decay of the actualized mental state, while (ii) the successive updates (formalized by σ_3) stabilize this state in one of the two representations . Let us emphasize again that we do not require the decay or update dynamics to be a genuine quantum process. Nevertheless, generalized quantum theory allows us to speak of complementarity in a well-defined manner.

The calculations for the Necker-Zeno model are the same as for the quantum Zeno effect, but the interpretation of the time scales is different. The probability that the mental representation has not changed after a time $t = N \cdot \Delta T$ due to N successive updates of the mental state separated by ΔT is given by

$$w_{\Delta T}(t) \;=\; \exp(-t/T) \quad , \tag{6}$$

with a mean dwell time T that satisfies:

$$T \cdot \Delta T \;=\; t_0^2 \quad . \tag{7}$$

We associate the following cognitive time scales with the parameters of the model (for more details see Atmanspacher *et al.*, 2004):

- ΔT is an internal update time for the mental state during observation of the Necker cube. We interpret this quantity as the interval between two successive stimuli that is necessary for a correct assignment of the sequence of their presentation. This so-called order threshold (Pöppel, 1997) is of the order of 25 – 70 ms. With smaller time intervals, the stimuli can still be distinguished but their sequence cannot be correctly determined.

- t_0 is the period of oscillations between the two states under the assumption that no updating observations take place. It is plausible to assume that the decay out of one state and the relaxation into the other occur on the same time scale. It can be related to the so-called P300 component in event-related potentials and is, thus, assumed to be of the order of 300 ms.
- T is the average time between successive switches of the mental state when the Necker cube is observed. It is usually characterized as roughly $T \approx 3\,\mathrm{s}$ and has large inter-individual differences (Brascamp *et al.*, 2005).

Relation (7) is clearly satisfied for these three cognitive time scales. Moreover, the predictive power of the model has been convincingly demonstrated with empirical results obtained under discontinuous stimulus presentation if it is possible to vary one of the time scales (t_0) as an independent variable and measure another one (T) as a function of t_0. Assuming that ΔT remains constant, Eq. (7) predicts a quadratic dependence for $T = T(t_0)$.

Under certain conditions, the time scale t_0 can be approximated by the off-time in discontinuous presentation, so it is indeed possible to test the model with experimental data. A comparison of observations by Kornmeier *et al.* (2007) with the predictions of the Necker-Zeno model is shown in Figure 4. The plotted symbols show observed values of T as a function of off-times. The solid curve represents a one-parameter fit of the Necker-Zeno prediction $T = t_0^2/\Delta T$ where $\Delta T \approx 70\,\mathrm{ms}$ gives the best results.

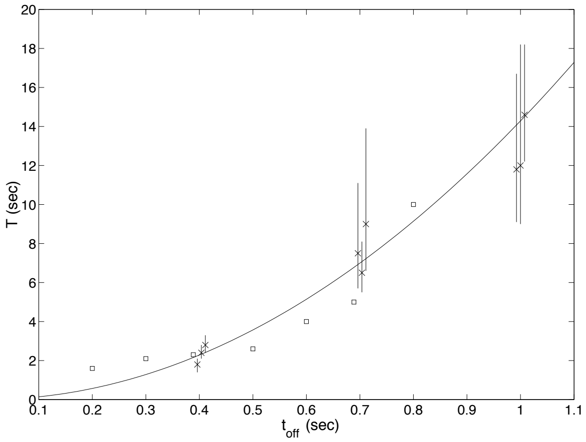


Fig. 4. Mean dwell time T as a function of the off-time t_0 in discontinuous presentation. The solid curve is a one-parameter fit (leading to $\Delta T = 70\,\mathrm{ms}$) from the prediction of the Necker-Zeno model under the assumption that the off-time can be identified with the decay time t_0 of the unperturbed system.

4.3 The Refined Necker-Zeno Model

While the Necker-Zeno model in the form just presented gives good predictions for the relation of the mean dwell time T to the other cognitive time scales ΔT and t_0, the predicted probability distribution of T differs significantly from experimental data. The cumulative probability that a change of the mental state has occurred up to time t is given by:

$$W(t) \;=\; 1 - w_{\Delta T}(t) \;=\; 1 - e^{-t/T} \quad , \quad T = t_0^2/\Delta T \quad . \tag{8}$$

From this we obtain the probability density $P(t)$ that a switch of the mental state occurs at time t:

$$P(t) \;=\; \frac{\mathrm{d}W(t)}{\mathrm{d}t} \;=\; \frac{e^{-t/T}}{T} \quad . \tag{9}$$

This probability density as well as the cumulative probability $W(t)$ are shown in Fig. 5.

A comparison with experimental data published by Brascamp *et al.* (2005) shows good agreement for times $t > 2\,\mathrm{s}$, but it reveals a completely different behavior for small t. The experimental probability density resembles a gamma distribution

$$P(t) \;\propto\; t^b\, e^{-\gamma t} \quad ,$$

with an exponent of $b \approx 5$, while the Necker-Zeno model according to Sec. 4.2 predicts a simple exponential decay. Similarly, the prediction for the cumulative probability from the Necker-Zeno model leads to linear behavior for small t while the experimental data rather exhibit power-law behavior with a large exponent.

Although the Necker-Zeno model in its original version was not intended to be valid for small values of t, it is nevertheless tempting to refine the model such that it provides the experimentally observed probability functions. The mathematical details of this refined version are published elsewhere (Atmanspacher *et al.*, 2008). We sketch the main results briefly.

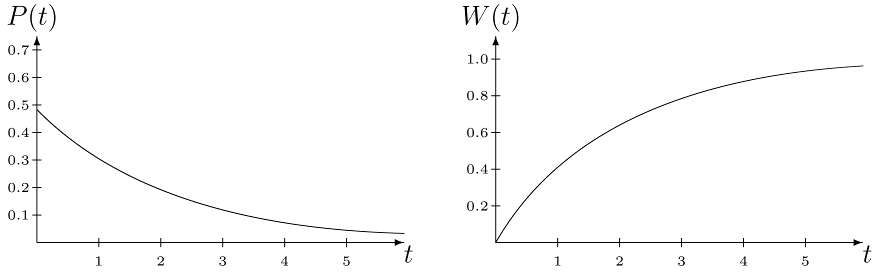


Fig. 5. The probability density $P(t)$ (left) for a switch of the mental state at time t and the cumulative probability $W(t)$ (right) that a switch occurred up to time t.

There are basically two possibilities to refine the original Necker-Zeno model for small values of t:

1. The parameter g, which determines the "decay velocity" of a mental state after the stimulus is turned off, is time dependent: $g \to f(t) \cdot g$.
2. The update intervals ΔT are time dependent: $\Delta T \to f(t) \cdot \Delta T$.

For large values of t the function $f(t)$ in both options approaches the constant value 1, such that the orignial Necker-Zeno model is recovered in this regime. For small values, however, $f(t)$ starts from zero with some power-law as shown in Fig. 6. Both a shorter update time for small t (case 2) as well as a slower decay for small t (case 1) can be interpreted as a form of increased attention. For more details concerning possible interpretations and applications see Atmanspacher *et al.* (2008) and Franck and Atmanspacher (2008).

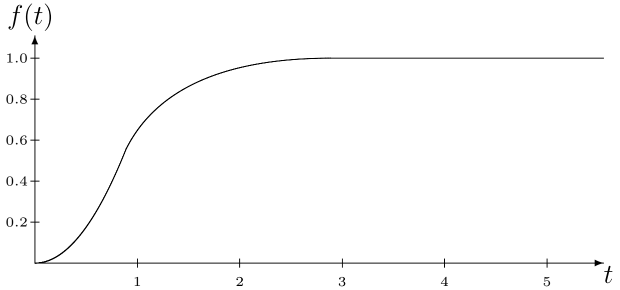


Fig. 6. The function $f(t)$ which determines the small-t behavior of either the parameter g or the update time ΔT in the refined Necker-Zeno model.

Despite the fact that in the probability distribution according to (2) only the product $g \cdot \Delta T$ enters, the physical interpretation of ΔT as a time interval leads to a different behavior of the probability distributions considered as a function of time. If for small values of t the function $f(t)$ starts with t^k, the behavior of $W(t)$ for small t is given by either t^k (for case 2, $\Delta T \to f(t) \cdot \Delta T$) or by t^{2k} (case 1, $g \to f(t) \cdot g$). Hence, the observed large values for the power law in the probabilities for small t are explained more naturally if g rather than ΔT is considered as time-dependent.

5 Temporal Bell Inequalities

In 1964, John Bell derived a set of inequalities which the expectation values of observables have to satisfy in any theory (i) that is local (i.e., any causal dependence respects the constraints given by Einstein's theory of relativity) and (ii) for which the results of a measurement are (at least in principle) already determined before the measurement is actually performed (Bell, 1966).

This second requirement was used by Einstein *et al.* (1935) in their famous EPR-argument as a definition of "elements of reality".

Experimental tests provided clear evidence that quantum mechanics violates Bell's inequalities (Aspect *et al.*, 1982a, 1982b). In particular, this implies that the statistical aspects of quantum mechanics cannot be explained by the introduction of local hidden variables. In general, it is assumed that in quantum mechanics the outcome of a measurement is not predetermined, even in principle, but rather the result of the measuring process itself.

Let Q_1, Q_2, Q_3, Q_4 be four different observables for which the result of a measurement can only assume one of the two values $+1$ or -1. In quantum mechanics, observables of this type are typically realized by measurements of polarizations of photons or by spin orientations of electrons. Let $E_{ij} = \langle Q_i Q_j \rangle$ be the expectation value for the (simultaneous) correlation function of Q_i and Q_j for the same system. One form of Bell's inequalities for this situation would be:

$$-2 \; \leq \; E_{12} + E_{23} + E_{34} - E_{41} \; \leq \; +2 \quad . \tag{10}$$

This inequality can be violated in quantum mechanics.

Bell's inequalities are expressed in terms of expectation values of two (or more) observables. The violation of Bell's inequalities in quantum mechanics involves the expectation values of (pairwise) non-commuting observables. Such observables cannot be measured simultaneously with arbitrarily high accuracy. In order to test the violation of Bell's inequalities in quantum mechanics, one makes use of particular correlations between two spatially separated systems which, however, are in an entangled state. Only under the assumption that "elements of reality" exist can one interpret the results as simultaneous correlation functions for one of the systems.

Obviously, the requirements for measuring a violation of Bell's inequalities as the key criterion for non-classical behavior are quite high. For a possible application to mental systems, the preparation of entangled states may be a particularly difficult problem. In addition, despite the fact that the quantum Zeno model provides the necessary non-commuting observables, the Necker-Zeno model for bistable perception has only one observable (σ_3) that serves to describe an "observation" of one of the two perspectives of the Necker cube. It is not clear if any of the other observables of the quantum Zeno model makes sense as an additional observable for the Necker-Zeno model. From this point of view it seems almost hopeless to test Bell's inequalities in the context of bistable perception.

But there may be an alternative option. In 1985, Leggett and Garg derived a set of inequalities which involve the expectation values of correlations of *one observable measured at different time instants* (Leggett and Garg, 1985). These so-called *temporal* Bell inequalities can be formulated in generalized quantum theory if the dynamics of a system does not commute with the observable (i.e., if the observable is not a constant of motion).

This is precisely the case for the Necker-Zeno model. Let $K(t_i, t_j) = \langle \sigma_3(t_i)\sigma_3(t_j)\rangle$ be the expectation value for a first "observation" of one of the two perspectives on the Necker cube at time t_j and a second one at time t_i. Then the following inequality should hold, if the mental state follows a "classical trajectory" (like the one shown in Fig. 2) with respect to the representation of the Necker cube:

$$|K(t_1, t_2) + K(t_2, t_3) + K(t_3, t_4) - K(t_1, t_4)| \leq 2 \quad . \tag{11}$$

If, on the other hand, during the periods of non-observation the mental state cannot always be described in terms of one of the two perspectives, this inequality can be violated. This might, for instance, be the case if the mental state is in a kind of superposition with respect to the two perspectives.[7]

For spin models, the largest violation of inequality (10) has been found to occur for measurements of the spin orientation along the angles $0°, 45°, 90°$ and $135°$. Similarly, one would expect the largest violation of temporal Bell inequalities in the framework of the Necker-Zeno model to occur at times t determined by the conditions $gt = 0, \pi/4, \pi/2, 3\pi/4$. For $t_0 = 300\,\mathrm{ms}$ this would correspond to measurements at $t = 0, 236, 471$ and $707\,\mathrm{ms}$. In order to avoid effects of observational updates, these values should be the off-times between brief stimulus presentations and observations.

This option to detect non-classical behavior in mental systems is as thrilling as challenging, but there is an important *caveat* to it. In the derivation of inequality (11) it is assumed that observations made on the same system do not influence each other. This is necessary for the determination of $K(t_i, t_j)$: The first measurement at t_j should have no effect on the result of the second measurement at t_i.[8] Such a requirement of "non-invasive measurements" might be difficult to realize for temporal Bell inequalities. On the other hand, it might be possible to estimate the degree of how much an observation of a mental state at time t_1 influences the observation of a mental state at time t_2. If this influence is smaller than the effect by which temporal Bell inequalities are violated, this could provide a terrific route toward evidence for non-classical behavior in mental systems.

---

[7] This gives an interesting twist to a question posed by Sudarshan (1983, p. 465): "Can we perceive a quantum system directly?" He speculates about a mode of awareness in which (p. 466) "sensations, feelings, and insights are not neatly categorized into chains of thoughts, nor is there a step-by-step development of a logical-legal argument-to-conclusion. Instead, patterns appear, interweave, coexist; and sequencing is made inoperative. Conclusion, premises, feelings, and insights coexist in a manner defying temporal order."

[8] This requirement corresponds to the locality requirement for the expectation values E_{ij} when the measurements are performed on different (but entangled) parts of a system: A measurement of Q_i at one part should have no influence on the measurement of Q_j at the other part.

6 Summary

The concept of complementarity has been defined in an axiomatic framework generalizing the quantum mechanical axioms for states and observables to systems involving invasive and thus, in general, non-commutative operations. In this framework, a novel approach to understand the bistable perception of ambiguous stimuli has been achieved, where the dynamics of the switch between different representations of a stimulus (e.g., the Necker cube) is complementary to the process of observation of these representations.

The corresponding Necker-Zeno model, referring to mental states and observables as well as their dynamics, is in agreement with experimental data for (1) the dwell time distributions (inverse reversal rates) in bistable perception and (2) the dependence of dwell times on off-times if stimuli are presented discontinuously. Finally, we have speculated about the possibility to formulate temporal Bell inequalities for this scenario. Their violation would imply evidence for fundamentally "non-classical" behavior in mental systems.

References

Aspect, A., Dalibard, J. and Roger, G. (1982a): Experimental test of Bell's inequalities using time-varying analyzers. *Physical Review Letters* **49**, 1804–1807.

Aspect, A., Grangier, P. and Roger, G. (1982b): Experimental realization of Einstein–Podolsky–Rosen–Bohm Gedankenexperiment. A new violation of Bell's inequalities. *Physical Review Letters* **49**, 91–94.

Atmanspacher, H., Bach, M., Filk, J., Kornmeier, J. and Römer, H. (2008): Cognitive time scales in a Necker-Zeno model for bistable perception. Manuscript available at `www.igpp.de/english/tda/pdf/neckerzeno.pdf`.

Atmanspacher, H., Filk, T. and Römer, H. (2004): Quantum Zeno features of bistable perception. *Biological Cybernetics* **90**, 33–40.

Atmanspacher, H., Filk, T. and Römer, H. (2006): Weak quantum theory: Formal framework and selected applications. In: Adenier, G., Khrennikov, A., and Nieuwenhuizen, T.M. (eds.), *Quantum Theory: Reconsideration of Foundations 3*. American Institute of Physics, New York, pp. 34–46.

Atmanspacher, H., Römer, H. and Walach, H. (2002): Weak quantum theory: Complementarity and entanglement in physics and beyond. *Foundations of Physics* **32**, 379–406.

Bell, J.S. (1966): On the problem of hidden variables in quantum mechanics. *Reviews of Modern Physics* **38**, 447–452.

Blake, R. and Logothetis, N.K. (2001): Visual competition. *Nature Reviews Neuroscience* **3**, 1–11.

Brascamp, J.W., Ee, R. van, Pestman, W.R. and Berg, A.V. van den (2005): Distributions of alternation rates in various forms of bistable perception. *Journal of Vision* **5**, 287–298.

Carter, O.L., Presti, D.E., Callistemon, C., Ungerer, Y., Liu, G.B. and Pettigrew, J.D. (2005): Meditation alters perceptual rivalry in Tibetan Buddhist monks. *Current Biology* **15**, R412–R413.

Einstein, A., Podolsky, B. and Rosen, N. (1935): Can quantum-mechanical description of physical reality be considered complete? *Physical Review* **47**, 777–780.

Franck, G. and Atmanspacher, H. (2008): A proposed relation between intensity of presence and duration of nowness. *This volume.*

James, W. (1890a): *The Principles of Psychology. Volume One.* Holt, New York.

James, W. (1890b): *The Principles of Psychology. Volume Two.* Holt, New York.

Kornmeier, J., Ehm, W., Bigalke, H. and Bach, M. (2007): Discontinuous presentation of ambiguous figures: How interstimulus-interval durations affect reversal dynamics and ERPs. *Psychophysiology* **44**, 552–560.

Kruse, P. and Stadler, M. (eds.) (1995): *Ambiguity in Mind and Nature: Multistable Cognitive Phenomena.* Springer, Berlin.

Leggett, A.J. and Garg, A. (1985): Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? *Physical Review Letters* **54**, 857–860.

Long, G.M. and Toppino, T.C. (2004): Enduring interest in perceptual ambiguity: Alternating views of reversible figures. *Psychological Bulletin* **130**, 748–768.

Mahler, G. (1994): Temporal Bell inequalities: A journey to the limits of "consistent histories". In: Atmanspacher, H. and Dalenoort, G.J. (eds.), *Inside Versus Outside. Endo- and Exo-Concepts of Observation and Knowledge in Physics, Philosophy and Cognitive Science.* Springer, Berlin, pp. 195–205.

Meng, M. and Tong, F. (2004): Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *Journal of Vision* **4**, 539–551.

Misra, B. and Sudarshan, E.C.G. (1977): The Zeno's paradox in quantum theory. *Journal of Mathematical Physics* **18**, 756–763.

Pöppel, E. (1997): A hierarchical model of temporal perception. *Trends in Cognitive Science* **1**, 56–61.

Sudarshan, E.C.G. (1983) Perception of quantum systems. In: Merwe, A. van der (ed.), *Old and New Questions in Physics, Cosmology, Philosophy, and Theoretical Biology.* Plenum, New York, pp. 457–467.

Process Ontology from Whitehead to Quantum Physics

Joachim Klose

Konrad-Adenauer Stiftung, Dresden, Germany, `Joachim.Klose@kas.de`

Summary. Although Alfred North Whitehead probably did not know much of the new quantum theory of Heisenberg, Schrödinger and Dirac, there seem to be deep similarities between his idea of process and the ideas of quantum theory. Both Whitehead's metaphysics and quantum theory are theories of observations: The realities which quantum theory deals with are based on observations by scientists who use the theory. And Whitehead's speculative cosmology is an expansion and generalization of the British empiricists' theory of perception.

Four leading ideas have determined the theoretical sciences in the 19th century: Atomicity, continuity, energy preservation and evolution. According to Whitehead, the challenge to science was not to introduce these concepts but to fuse them together and expand their application. Therefore, the cell theory and Pasteur's work were more revolutionary for him than the achievement of Dalton's nuclear theory, "for they introduced the notion of organism into the world of minute beings. . . . The doctrine of evolution has to do with the emergence of novel organisms as the outcome of chance" (Whitehead, 1925, pp. 146–147).

Up until now, neither individual experiences nor the natural sciences gave reason to believe in invariable subjects. On the contrary, the whole being of reality has been in a process of becoming and passing. "On the organic theory, the only endurances are structures of activity, and the structures are evolved" (Whitehead, 1925, p. 158). Whitehead's speculative cosmology is based on the results of the theory of evolution. However, he tries to integrate all experiences of reality. Placing the concept of "actual occasions" in the center of his philosophy of organism, he succeeds in resolving handed-down contrasts within a common framework. The world is made of 'actual occasions', each of which arises from potentialities created by prior actual occasions. Actual occasions are "happenings", each of which comes into being and then perishes, only to be replaced by a successor. These experience-like "happenings" are the basic realities of nature.

Similarly, Heisenberg said that what really happens in a quantum process is the emergence of an "actual" from potentialities created by prior actualities. In the orthodox Copenhagen interpretation of quantum theory, the actual things to which the theory refers are increments in "our knowledge". These increments are experiential events. The particles of classical physics lose their fundamental status: They dissolve into diffuse clouds of possibilities. At each stage of the unfolding of nature,

the complete cloud of possibilities acts like the potentiality for the occurrence of a next increment in knowledge, which can radically change the cloud of possibilities and potentialities for later increments in knowledge.

A philosophy founded on causality and teleology as basic descriptions of reality must dissolve the distinctions between inside and outside, consciousness and matter, object and subject. To achieve this purpose, Whitehead's philosophy of organism offers a starting point. Therefore, I would like to introduce his philosophy and compare its results with interpretations of quantum theory. Here, it will be interesting to take a look at Henry Stapp's theory of consciousness, which is based on quantum theory. He argues that reality is created by consciousness, as consciousness causes the collapse of the wave function that in turn causes reality to "occur". Stapp claims that Whitehead's metaphysics is incompatible with quantum theory by virtue of Bell's theorem and needs to be modified. I disagree with this conclusion because Stapp did not properly take into account Whitehead's theory of prehension.

1 Introduction

There have been countless discussions about the implications of physics, especially quantum physics, for various issues of human understanding. These issues include time, consciousness, and freedom (Griffin, 2005).

- Regarding *time*, it has been argued that modern physics shows time as we experience it – with its distinctions between past, future and present – to be ultimately unreal.
- Regarding *consciousness*, it is thought that any philosophy of mind, to be compatible with modern physics, must regard conscious experience as a by-product of the brain's subatomic particles.
- Regarding *freedom*, it is thought that any understanding of reality based on modern physics must rule out the possibility that our decisions truly involve self-determination.

In light of these supposed implications, it is widely assumed that a worldview that takes physics seriously necessarily contravenes the worldview of ordinary human understanding. In reality, none of these implications must follow from physics *per se*. These are always interpretations from a particular philosophical perspective. Physics as interpreted by Whitehead's philosophy rejects all three implications. They are examples of what he calls "the fallacy of misplaced concreteness", meaning the "error of mistaking the abstract for the concrete" (Whitehead, 1925, pp. 74–75).

By characterizing the basic ideas of sciences and their consequences for philosophy, Whitehead wants to unify different views of nature and to overcome the dualistic tradition of Cartesianism in modernity (Whitehead, 1925). One can summarize his effort against dualisms of three versions (Wiehl, 1998):

Ontological dualism: This denotes the absolute difference between an infinite and a limited substantiality.

Ontic dualism: This denotes the absolute difference between the physical and the spiritual being.

Gnoseological dualism: This denotes the absolute difference between two kinds of knowledge, different in nature, between rational grounds and grounds of experience.

It is interesting that Whitehead's starting point in the analysis of the ideas of the 19th century resembles that of Friedrich Engels. Both selected a nearly identical group of scientific advances which they saw as the deciding factors in the transition from Newtonian to modern science: Atomicity, continuity, energy preservation and evolution. In addition, Whitehead's philosophy of experience resembles dialectical epistemology in stressing the role of negatives (Wiehl, 2000, p. 40). However, instead of representing a dialectical materialism, he arrives at completely different conclusions. He asks himself whether we can "define an organism without recurrence to the concept of matter in simple location" (Whitehead, 1925, p. 149) and radically rejects every type of materialism.

In terms of experience, language and logic, the teaching of a substance forming the basis of all things seems to be the most natural way to look at last things. In modernity, however, the originally "logical" substance-quality pattern has been raised to the basic structure of reality. The result is that relations between things can no longer be taken into account (Whitehead, 1978, p. 79). Whitehead rejects:

1. a substance as a static substratum (Christian, 1959, p. 108) because
   a) we experience a variable world, and
   b) the natural sciences are becoming smaller and smaller "particle unities" (Whitehead, 1978, pp. 78–79);
2. "the fallacy of simple location" because the objects of the world exist neither in isolation nor independently of one another.

Neither individual experiences nor the natural sciences give reason to believe in invariable subjects; on the contrary, the whole being of reality is in a *process* of becoming and passing. Whitehead suggests seeing reality as analogous to an organism (Whitehead, 1925, p. 159). His speculative cosmology in "Process and Reality" (Whitehead, 1978) is the logical construction of a philosophy of organism based on the results of the theory of evolution. However, he tries to integrate all experiences of reality. For this reason, he criticizes Darwinism, which completely excludes creativity.

2 Bifurcations

The sciences are not concerned with epistemological matters but rather with a coherent explanation of nature. This fact leads to the bifurcation of reality. Whitehead categorically rejects:

- the distinction between events of nature and events as they are formulated in scientific theories, and
- the distinction between events of nature as they exist by themselves and as they appear to us.

The first concept maintains a purely conceptual existence of physical entities such as atoms and electrons. On the one hand, there are phenomena, and on the other hand, logical terms of scientific formulae. For Whitehead, scientific concepts are derived from nature by way of logical abstraction. He argues against the bifurcation of reality into the mathematical world and the apparent world. Concepts, as far as they are true, refer directly to facts of reality.

The second bifurcation is a consequence of the first. It appears between sensory perception and reality itself and results in the banishing of the observer from nature. The observer can have knowledge only of his sensory impressions, not of the objects which produced them. The knowledge of reality now requires a theory since there is a rationally unbridgeable gap between the purely geometrical concepts of motions of particles in space and the psychological realities of conscious sensations, feelings, and ideas. If the material substances are only in space, then a material substance can act only upon material substances – not upon a mental substance. Gottfried Wilhelm Leibniz concluded (Schilpp, 1941, p. 179),

> "either the material substances must be brought within the space in the field of awareness of the mental substance ... or the mental substance must be defined in terms of the material substances."

For Leibniz the latter is impossible because he would then be confronted with Locke's problem.[1] This is why he developed his monads.

3 Perception

To avoid these bifurcations, the origin of every possible knowledge must be considered. Whitehead regards this origin within everyone's daily experiences and addresses directly the British empiricists' starting point:

1. Every experience has its origin in perceptions.
2. The primary ideas of perception join secondary ideas deduced by reflection in order to put the sense data into an order.
3. In addition to these two starting points of the British empiricists, Whitehead integrates psychic impressions such as emotions, beauty, love and satisfaction.

---

[1] Historically, after separating the realm of apparent nature from that of its physical description, John Locke asked how both realms could be connected. Isaac Newton developed a kinetic theory of atoms, but he did not explain how unperceivable atoms in absolute space and time are connected with our space-time experiences.

If Whitehead would restrict himself to the British empiricists' theory of perception, he would be subjected to the false conclusion that reality is constituted out of static and isolated substances. Because he does not, he can do justice to relations in sensory perception due to the broadening of the theory of perception through the mode of "causal efficacy".

Perceptions are normally described in "presentational immediacy". This mode of perception presents the spatial relationships between the perceiver and sense data, even while temporal aspects are ignored. Perceptions in presentational immediacy are preferred compared to causal efficacy because they are directed by attention. Attention is comprised of a teleological and a temporal aspect. The analysis of past data directs the attention to the emergence of future data. However, the analysis of past data is no longer part of presentational immediacy but rather of causal efficacy. Attention is the cut between presentational immediacy and causal efficacy. All scientific observations are made in the perceptive mode of presentational immediacy (Whitehead, 1978, p. 169). However, physical theories refer exclusively to causal efficacy.

If all knowledge is traced back to perception at one moment, one cannot have empirical knowledge of relations nor of the continuum of reality. Contrary to David Hume and Immanuel Kant, Whitehead finds evidence for causal connections and temporal continuity in sensory perception. He asserts that one can perceive them directly in the mode of causal efficacy, tacitly assuming the experience of temporal and spatial extension.

Temporally adjacent events are perceived directly in a temporal window of perception: the "specious present" (according to William James). It is perceivable that later events confirm earlier ones. We have knowledge of an extensive continuum of reality because of our perception of space-time relations. The specious present contains not only immediately observed events; it also includes the immediate past. The presence of immediately past events shows that present and future events have to confirm earlier events in the same way that immediately past events had to confirm events in the even more distant past. Causality in Whitehead's philosophy means that we never perceive a series of events alone; later events must emerge from earlier events in the specious present. Perceptions in causal efficacy contain the temporal aspects of the process of reality.

Sensory perception takes place only in the complex mode of "symbolic reference" connecting the two pure modes. As a result, perception in symbolic reference causes errors and misinterpretations. Symbolic reference is an active synthetic element of the perceiver, producing emotions, convictions and beliefs concerning other elements of reality.

4 Time

Within each period of his philosophical development, Whitehead expresses that space and time do not exist independently. Space-time cannot be considered a self-subsistent entity. It is an abstraction whose explanation requires reference to that from which it has been abstracted. In Whitehead's natural philosophy, the real world is an extended, continuously flowing process. Later, in his metaphysical period, space and time were seen as abstractions from extended events and are to be experienced empirically.

4.1 Time in Whitehead's Natural Philosophy (1914-1925)

An entity is an abstraction from the totality of the continuously flowing process of reality. Temporally extended events do not exist independently. What scientists accept as elements or parts of the whole are actually abstractions. In reality, the elements only exist and have meaning by virtue of the whole and *vice versa*. Therefore, time does not have any reality in nature but is the property of a perceiver. Reality is characterized by an extensive space-time continuum. Events in nature do not have any reality independent of a consciousness and do not have definite temporal extensions. Time relations are an expression of an ordering relation of a perceiver. Space-time is nothing other than a system for the combining of assemblages into unities. Physical time only deals with certain formal, relational aspects of our changing human experience. Relative to other abstractions, space and time offer a comparatively simple structure, which is suitable as a basis for objective distinctions in reality.

During the specious present one perceives a unit already separated into its parts by the activity of the perceiver. The parts entertain certain characteristics, of which time and space are examples. The common structure of space-time conforms to the uniform experiences of sensory perception. But it is not clear how one can proceed from individual experiences to a uniform space-time structure. Whitehead confesses that what he has termed the "uniformity of the texture of experience" is a mere illusion. This uniformity does not belong to the immediate relations of the crude data of experience but, rather, is the result of substituting more refined logical entities. We are not directly aware of a smoothly running world.

4.2 The Epochal Theory of Time (after 1925)

The transition from momentary events to extended events is not only initiated by the knowledge that perception takes place in the specious present and that causal interactions are directly perceivable. It is also a result of logical difficulties within physical theories and metaphysical outlines. Physical descriptions of dynamic processes like momentum, velocity and tension, and the descriptions of simple physical structures like atoms or biological organisms presuppose the existence of temporal events. In addition, becoming is only possible if reality is constituted out of temporal, atomic events. Becoming and continuity are incompatible (Zenon of Elea). Whitehead shows that momentary events can be deduced out of extended events by means of the method of extensive abstraction, one of his central ideas. All these points forced him to conclude that reality is not founded on momentary events but rather on spatiotemporally extended events.

Despite the fact that Whitehead probably never became acquainted with the post-1924 development of quantum theory, first results motivated him to transfer the new knowledge of philosophy and psychology to all events of reality. In particular, Bohr's model of the atom (1913) and de Broglie's wave theory (1924) resulted in a critical examination of his natural-philosophical starting point. From that point on, the particles of reality were no longer material, static forms but rather spatiotemporally extended events. The change from materialism to Whitehead's organic realism is characterized by the displacement of the notion of static stuff by the notion of fluid energy. Whitehead got his inspiration from scientific discoveries, without necessarily going into their specific formalism. His doctrine of the epochal character of time depends on the analysis of the intrinsic character of an event, considered to be the most concrete, finite entity, which he calls the "actual occasion".

In the epochal theory of time, Whitehead unifies four different time aspects to be found in the experience of an actual occasion. There are two internal and two external aspects. The internal time aspects are the passage of thought (becoming and perishing, retentions), and the experience of extension (unlimited act, inner time consciousness, retentions and protentions). The external time aspects are the potential physical time (extensive continuum), and the actual physical time (passage of nature, becoming and perishing). The experience of extension corresponds to potential physical time; the passage of mind corresponds to the passage of nature. The physical concept of time unifies the experience of an extensive continuum and the perception of concrete, actual occasions. It unifies the discontinuity and continuity of the external world into one concept.

5 Actual Entities

The assigning of the different time aspects to the final units of reality becomes possible through the transformation of the concept of momentary events into actual occasions.[2] While in Whitehead's natural philosophy events still depend on the activity of a perceiver, actual occasions are in his metaphysics the final units of reality (Whitehead, 1978, p. 75). They are the real things of the world and have their own being. They are not momentary cuts through reality but rather forms which have the properties of spatiotemporal extension and creativity.

The adjective "actual" rejects every attempt to find a reality behind actual entities (Whitehead, 1978, p. 75). An actual entity is limited in terms of space

---

[2] Whitehead uses the notion of "actual occasions" interchangeably with "actual entities".

and time and, in comparison to other actual entities, owns a defined space-time position (Whitehead, 1978, p. 73). It follows that an actual entity neither moves (Whitehead, 1978, p. 77) nor changes! Entities appear and disappear like the ideas in the stream of ideas in our mind (Whitehead, 1978, p. 141).

Every actual occasion is a spatiotemporal unit possessing an indivisible volume and time quantum, which cannot be disassembled without being destroyed (Whitehead, 1978, p. 219). Actual occasions express the uniform space-time structure of the universe because their external relations fit them into superordinate actual occasions, and their internal relations, the coordinate divisibility, divide them into subordinate actual occasions. The spatiotemporal extensive continuum is the general structure to which all actual occasions must conform. Actual entities, whose unity can be dissolved into subordinated actual entities, are called *Nexus.* The usual things, like trees, houses and cars, are all Nexūs (Whitehead, 1978, p. 56). Nexūs take into account the unity of *contemporary* events which are not causally tied together. If a Nexus owns an ordinal degree, Whitehead calls it "society" (Whitehead, 1978, pp. 89–90). "A society is a sequence, or more generally, a pattern of occasions which . . . give rise to the impression of objects existing self-identical in time" (Hättich, 2004, p. 101).

6 Prehension

The content of an actual entity is constituted only by perceptions, like the contents of Locke's "idea". The "perceiving" actual entity is connected with other entities by perceptions. Whitehead's philosophy of organism is a generalization and extension of his theory of perception. However, perception is not limited to sensory perception but refers to every kind of causal influences.

"Prehension" is a short form of "apprehension", which indicates "recognition" and does not mean that the perceived has to be present. However, it presupposes consciousness. Consciousness belongs only to a few highly-developed organisms. Whitehead's concept of perception should describe universals and should also contain "unaware recognition". Therefore, he introduced the concept "prehension". Every entity which is prehended as a unity is an actual entity. "God is an actual entity, and so is the most trivial puff of existence in far-off empty space" (Whitehead, 1978, p. 18). Actual entities are not only microcosmic entities as is often maintained (Sherburne, 1966, p. 205).[3] For Whitehead, the whole universe as well as just a single atom are actual entities.

An actual entity is linked with *every* other actual entity of the universe by means of prehensions (Whitehead, 1978, p. 41). Although all actual entities of the world are prehended, not every actual entity contributes to the new

---

[3] Abner Shimony's paper on "quantum physics and the philosophy of Whitehead" is based on this misconception (Shimony, 1965).

actual entity. Otherwise, all actual entities would be the same and therefore indistinguishable. The becoming actual entity selects all "positive" prehensions for its construction. They are called "sensations" or feelings. An actual entity "feels" the contributions of other actual entities and integrates them into its construction (Whitehead, 1978, pp. 56–57).

There is a significant difference between perception, which is causally influenced by perceived objects, and prehension, which means a coming together of different parts of reality. The latter could also mean a going together of very distant events. Thus, there is also a strand in Whitehead's metaphysics discussing parapsychological phenomena, especially telepathy (Whitehead, 1978, p. 308; Griffin, 1982). For Whitehead, "physical science maintains its denial of 'action at a distance,' the safer guess is that direct objectification is practically negligible except for contiguous occasions" (Whitehead, 1978, p. 308).

In this respect, I believe that Henry Stapp did not sufficiently take into account Whitehead's discussion of prehensions when he claimed that Whitehead's system of metaphysics is incompatible with quantum theory due to Bell's theorem. In contrast, Bell's theorem could be used to support process philosophy (Klose, 2002, pp. 355–357). Each event does prehend all of creation, not only those events found in its backward light-cone, as Stapp (1977, p. 315) predicated. The unity of the world would be destroyed if each event would prehend only its own actual world (Stapp, 1979, p. 21).

A theory of perception connects causally past events with present ones. But the theory of prehension changes the perspective. It describes the development of reality from present to future. Therefore, a growing actual entity is not the perceiving subject in the process of prehension. The perceiving subject does not exist before the perceived events and is not their contemporary. This would mean a new formulation of a concept of substance, of a basis bearing the phenomena. *Vice versa*, the perceived events are temporal before the objectifying actual entity. Prehensions reach into the future like tentacles. They grow together into a new unity.

However, this process does not take place locally and aimlessly. It is accompanied by an ideal, the subjective aim. Actual entities lead their incremental process. They present themselves as the aim of this process. In this respect, they are both subject and superject in one event, the superject being the decisive element in the process. Whitehead generalizes the structure of perception of a consciousness. He ascribes this structure to nature as a basic structure of reality. Nature does not appear anymore as coexisting, separated particles of matter but rather as a network of organically interconnected entities.

7 Creativity

Every future entity means a coming together of all available elements of reality. The fact that every entity of reality tends to unification and to higher complexity is an empirical fact (Whitehead, 1929, p. 89). The internal, motivating force of the reality process are creative processes of becoming, which Whitehead calls “concrescence” (Whitehead, 1978, pp. 21–22). The philosophy of organism is based on the generalization of the concept of force (Rapp, 1986, p. 82). One constitutive quality possessed by all entities is creativity. It is the impetus of progress to new units of reality.

The standard (neo-Darwinian) theory of evolution does not explain why evolution as a whole has led towards ever more complicated life forms. At variance with the general laws of physics, which postulate that there will be an equal distribution of energy and decay throughout the universe, we know that processes leading to higher organization forms exist (Whitehead, 1929, p. 24). We know from our observation of human and animal experiences that purposes are immediate components of the constitution of living beings (Whitehead, 1929, p. 13). Physiology and physics, which describe reality only in terms of active causes, ignore these experiences. Therefore, their theories are not adequate descriptions of reality as a whole.

An adequate description of the universe has to contain aspects of both efficient and final causation. For this reason, we must not describe nature only in terms that ignore one side of reality. The only kind of entities observable in nature are living organisms, which unify final and efficient causation. It is more reasonable to transfer the concept frame of living organisms to all phenomena of reality than the reverse (Whitehead, 1938, p. 211). Whitehead’s philosophy of organism attributes the double character of efficient and final causation to the final things of the universe.

Whitehead identifies the energetic activity of physical entities with the emotional intensity which can be perceived in the life of biological nature (Whitehead, 1938, p. 231). All entities of reality are “living beings”. Neither the nature of physical entities nor life can be understood independently of each other. Life implies self-preservation, creative activity, and teleological aim.

The opposites of “efficient causation–final causation”, “decay–pursuit of higher complexity” and “body–mind” are unified in the concept of life. All events of reality live if they comprise these tensions. According to neo-Darwinian evolution, primordial physical events enter into mental events and cause them. According to the philosophy of organism, the reverse is basic. It takes back the grounds of mental events by using physical ones. Every event possesses (a) mental and (b) physical experiences (Whitehead, 1926, p. 118):

- Mental experiences are experiences of defined forms (universals, eternal objects), regardless of their concrete determination of being.
- Physical experiences are conservations of facts given to the event by its constitution of being.

An actual occasion is the product of the interplay of the physical with the mental pole. The physical pole is extended over the whole space-time continuum and can be divided. In contrast, the mental pole does not share in the divisibility of the physical pole. The mental pole has its equivalent in a thought (of mind). It is an act of attention with the duration of the specious present. Passage of mind is confronted with the experience of an unlimited temporal act in the internal time concept.

8 Teleology

Present actual entities do not anticipate their future determination but rather their subjective aim. If an actual entity always and unavoidably reaches its subjective aim, all future events would be determined by present ones and *vice versa*. This is not the case. The subjective aim is a future aim of a present development envisaged by the becoming actual entity. As the entity is an eternal object, it is the vision of a future state, which influences the way it develops into this state. The vision influences the actual entity in its decision but is not the final determination (Whitehead, 1933, p. 249):

> "In the formation of each occasion of actuality the swing over from re-enaction to anticipation is due to the intervening touch of mentality. Whether the ideas thus introduced by the novel conceptual prehensions be old or new, they have this decisive result, that the occasion arises as an effect facing its past and ends as a cause facing its future. In between there lies the teleology of the universe."

The difference between the present state of the development process and the subjective aim is the excitement pushing an actual entity forward to higher states of development. Its "appetite" for completion will have "evaporated" if a state of satisfaction is reached. An actual entity reaches fulfilment if the difference between the subjective aim and the satisfaction has become negligible. The process is finished at a certain state of convergence. One has to understand this approximation process as a process of fulfilling an ideal.

It could be concluded from the pursuit of the subjective aim that there has to be something within which the subjective aim is present. Something exists that moves towards this aim. This thought puts a subject under the reality process envisioning the subjective aim and is the medium of the process of development. This means that there has to be a substrate of changes in reality which contradicts Whitehead's intention. Where does one find the origin of the subjective aim in the concrescence process of an actual entity ?

Whitehead denies the "intentions" of past actual entities. They have passed away and do not possess a transition to future aims. A growing actual entity perceives the subjective aim as a date of the actual world. The subjective aim as a date is contemporaneous with past actual entities; as a purpose, it is neither a cause nor an effect. The "'moving' finis in the final nexus is the interpretation of the purpose as a cause." (Löw, 1980, p. 292) The subjective aim determines which prehension delivers a positive contribution to the growing actual entity. According to Whitehead, the subjective aim is made available by God. This is a crucial point. From where does mentality get a future picture? Are there experiences or concepts which are not reducible to the observed nature?

One can only distinguish between subjective aim and satisfaction if the process of concrescence is limited in time. Whitehead (1978, p. 19) took for granted that teleology assumes temporal atomicity, and that temporal atomicity is only possible in a state of reality which is teleological. In a cosmology with a continuous concept of time, real becoming is impossible – there are only changes which are transformations from one state into another. However, a physical process, which is teleologically constituted, assumes an aim of development for single entities.

9 Transmission and Concrescence

Process philosophy differs significantly from classical philosophical drafts in its dynamically oriented conceptual design of reality. Dynamic processes can be considered from an internal and from an external perspective (Whitehead, 1978, pp. 51–52). The internal process is the process of concrescence; it makes up the essence of actual entities and is teleologically structured. The external process characterizes the progress from actual entity to actual entity and describes changes within societies of actual entities. It is characterized by causality and conformity (Whitehead, 1978, p. 210).

Reality is the common presentation of two kinds of processes: Concrescence and transmission. The transmission process concerns the steady progress made by atomic unities of reality from the past to the future. This process is described by the theory of evolution. Transmission is a process of concrescence processes, and concrescence is a process of transmission processes. Reality is a process of processes. Every actual process contains a huge number of interlocking actual processes. The whole universe is a single process as well as an infinite complex of processes.

10 Quantum Theory

Whitehead was clearly influenced by the very early development of quantum theory, so one might expect similarities between quantum theory and his process philosophy. In particular, the properties of an actual occasion and a quantum event are quite similar. It appears that the collapse of the quantum state is the atemporal process that corresponds to an actual entity, and the elementary quantum event corresponds to what Whitehead called "the satisfaction of an actual entity" (Malin, 2006).

There is another parallel concerning the conception of a classical trajectory. It is a consequence of Heisenberg's uncertainty relations that a quantum particle cannot have a definite position in space and a definite momentum at the same time. Hättich (2004, p. 100) states:

> "Consequently, quantum particles cannot posses continuous trajectories because this would obviously force them to possess a definite position *and* a definite momentum at each time of their existence."

The experimental results one gets from a bubble-chamber experiment look like the spatiotemporally continuous trajectory of a classical particle. Again Hättich (2004, p. 100):

> "But under closer inspection it turns out that this 'continuous' trajectory is merely a succession of discrete, i.e. spatiotemporally non-overlapping, events."

This description of a trajectory is in accordance with Whitehead's concept of a society.

But in how far can Whitehead's metaphysics provide an ontological basis for quantum theory ? For Einstein, a theory always represents an extrapolation beyond what we can know (Haag, 2004, p. 54). Although Shimony (1965) concealed the usefulness of Whitehead's metaphysics for an interpreting system of quantum theory, lately some articles and books have been published on this subject (e.g. Burger, 1965; Griffin, 1982; Stapp, 1993; Shimony, 1993a, 1993b; Eastman and Keeton, 2004). There are strong endorsements of process philosophy, and striking parallels to Whitehead's formulations.

The "Copenhagen" quantum theory was formulated as a set of practical rules for making predictions about what human observers would observe under certain well-defined conditions.[4] This pragmatic view "is essentially subjective and epistemological, because the basic reality of the theory is 'our knowledge' " (Stapp, 2001a, p. 2). It contains in itself no definitive criterion of completeness. However, it is guided by two basic principles (Stapp, 1979, p. 9): "The final theory should be comprehensive and unified." In this regard, the Copenhagen formulation includes an awkward feature: Human observers are excluded from the system. The theory is based on a bifurcation of the physical world into observer and observed. This situation is dissatisfying for someone who seeks a rationally and dynamically coherent understanding of what is actually going on. Because measuring devices and human bodies are made up of atoms, one expects that the laws of quantum theory, if universal, ought to work for these physical systems, too.

Two choices enter into the determination of what happens in quantum theory in general and in quantum measurement in particular:

1. the choice of questions which are posed upon nature, and
2. the choices of the answer of nature to the chosen question.

Quantum theory gives statistical predictions for point (2). But the question in (1) is chosen by the experimenter. The exclusion of the experimenter from the

---

[4] This summary follows Stapp's (2004) ideas.

system being investigated is fixed by the "orthodox" quantum theory devised by von Neumann and Wigner. Von Neumann showed that the observed event in the external world is directly linked to the brain of the observer of that event. The observed system (process 2) is described in terms of quantum mathematics, the observing system (process 1) in terms of human experiences. Due to the fact that it makes no practical difference which of the various placements of the dividing line between the two systems one uses – the placing of the border is a matter of expedience (Haag, 2004, p. 54) – von Neumann put all parts of nature composed of atomic constituents on the side described in terms of the quantum mathematics and only the consciousness of the observer outside of the mathematically described world. In von Neumann's formulation, the whole world is treated as a quantum system.

Because von Neumann's theory is built on the Newtonian concept of an instant of time, it was elevated by Tomonaga and Schwinger to a form compatible with the physical requirements of the theory of relativity. In their relativistic quantum field theory, the Tomonaga-Schwinger surface σ does not differ significantly from the constant-time surfaces of Newtonian physics. Contrary to the theory of relativity, there is a preferred sequence of instantaneous "nows". Direct changes of a part of the surface σ cause *indirect* changes along the rest of the surface due to quantum entanglements. According to Stapp, these indirect changes produce the 'faster-than-light' effects, and elsewhere Stapp (2001b, p. 10) says:

> "Thus quantum theory reverts, at a certain deep ontological level, to the Newtonian idea of instantaneous action at a distance, while maintaining all of the empirical demands of the theory of relativity."

Nonetheless, there must be a dynamic connection between mind and brain: The mind of the observer is obviously connected to what is going on in his brain, and his choice of which question to put to nature influences his brain in ways controlled in principle by quantum laws. Asking a question about something is closely connected to focusing one's attention on it. Due to Stapp, this connection can be found via the quantum Zeno effect, which shows how the choice and timing of questions can influence the course of events in the probed system. Physical principles do not specify which questions are posed to nature. This opens the logical possibility that our conscious thoughts could be entering into the mind-brain dynamics in a way reducible neither to purely mechanical effects governed by the Schrödinger equation of motion nor to the random effects of nature's choices of outcomes.

In general, our thoughts issue commands to "attend" to certain questions in the future. These directives supply the missing component of the quantum dynamics: They pose the particular questions that are put to nature. The point is that the occurrence of a conscious thought associated with a quantum system is supposed to cause a reduction of the state of that system to the reduced state. Since the question to be posed is supposed to be an experience, it would appear that it really ought to be part of the mental, rather than physical, side of the mind-brain dynamics. Quantum theory has a *lacuna* that can very naturally be filled in such a way as to allow our thoughts to exercise real, though not absolute, control over the mechanical aspects of mind-brain dynamics.

11 Process Philosophy and Quantum Ontology

"The natural *ontology* for quantum theory ... has close similarities to key aspects of Whitehead's process *ontology*" (Atmanspacher, 2006, p. 71). Both are theories of perception. Whitehead tells us that it is equally possible to arrive at his organic conception of the world from psychology on the one side and from mathematical physics on the other (McHenry, 2002, p. 168). Otherwise, "quantum theory gives us a mathematical model, not of an independent reality, but of our perception of reality" (Hartshorne, 1977, p. 189). Both are interpreting systems of nature and share the same intention.

On the other side, all Whiteheadian-inspired physicists have in mind a discussion of "a modified philosophy of organism, which would preserve Whitehead's essential ideas while according with the discoveries of modern physics" (Malin, 2002, p. 172). There seem to be great differences deeply rooted in the concept of time. Spatially separated parts of reality must be related in some way that goes beyond the familiar idea that causal connections propagate only into the forward light-cone. Quantum events behave as a unified system: "What *you do to it* in one place can influence how it will react to a *simultaneous* probing far away" (Stapp, 1993, p. 30).

Whitehead has been blamed for having only a causal theory of perception, with which he cannot account for contemporary events (Stapp, 1979, p. 2). Actually, Whitehead introduces three different concepts of contemporaneity: Contemporaneity, simultaneity, and instantaneity. "An instantaneous space is static, being related to the static nature at an instant" (Whitehead, 1920, p. 117). "Actual entities are called 'contemporary' when neither belongs to the 'given' actual world defined by the other" (Whitehead, 1978, p. 66). This concept covers all events in the light cone. But simultaneity includes all contiguous events of prehension. These events need not be causally connected. Two electrons very distant from one another are also contiguous by means of gravity. Prehensions grow together to new actual occasions if they fit to each other, i.e., if they pass in coherence.

Process philosophy can cover the results of Bell and Tomonaga-Schwinger that the available information about a system can be effected by a far-away observation (Stapp, 2001a). For Whitehead, the available information about the (far-away) system which is disturbed by the (nearby) measurement and the nearby system are one actual occasion. There is no need to modify process philosophy at this point. On the contrary, it is actually a release that we have physical as well as philosophical reason to dismiss the idea of mutually independent events (Hartshorne, 1977, p. 185).

Einstein adopted the absence of absolute motion as one of the key postulates of the special theory of relativity. “This resulted in the ontology that the phenomenon of time was essentially an inseparable aspect to space itself; that reality was an unchanging piece of geometry” (Cahill, 2005, pp. 6–7). In this ontology there is no notion of change and becoming nor of the experiential aspects of time. Space itself, in conjunction with the sensitized detector, has some real role in the measurement procedure. Space turns out to be a dynamic system, not some passive piece of geometry.

Parisi and Wu discovered that a formalism of stochastic quantization underlies the functional formalism of Dirac and Feynman (Cahill, 2005, pp. 6–7). Stochastic iterative systems have essentially time-like properties. Why not abandon the static scheme underlying space-time, upon which quantum field theory is constructed, and keep only the stochastic iterative process ? Time would no longer be modelled by some fundamentally different system, such as by geometry, but by a time-like process itself. A stochastic iteration model contains no notion of space and matter. It is very similar to stochastic neural networks. If this model of reality proves to be successful, then one could adopt the ontology that reality is *mind-like*, as Leibniz and Whitehead suggested. “Because it involves a modelling of time which matches its experiential properties this radical new modelling of reality is called *process physics*” (Cahill, 2005, p. 11).

Does Whitehead’s ontology contain an inconsistency due to the fact that the principle of separateness of all realized regions will generally not be satisfied in his causally local and separable ontology (Hättich, 2004, p. 249) ? This would be true if his metaphysics were traced back only to the theory of relativity, if one did not take into account that his ideas originate from a psycho-philosophical discussion, that his theory of prehension connects all occasions of the contemporary world, and that the concrescence process selects positive prehensions. If one concluded that, then either the causal independence of simultaneous occasions or the distinctness of their concrescence processes would have to be abandoned in order to secure the separateness of all realized regions, and one would have to answer two questions: What does causality mean ? Likewise, what does separateness mean ?

In the words of Hartshorne (1977, p. 188):

> “Causality is merely the way in which each instance of freedom takes into account the previous instances, as each of our experience refers back through memory to our own past and through perception to the world’s past.”

According to quantum thinking and process philosophy there is no backward-in-time causation. Rather (Stapp, 1977, p. 321),

> “the basic properties of relativistic quantum theory emerge . . . from a logically simple model of reality. In this model there is a fundamental creative process by discrete steps. Each step is a creative act or event. Each event is associated with a definitive spacetime location. The fundamental process is

> not local in character, but it generates local spacetime patterns that have mathematical forms amenable to scientific studies."

And, again, Hartshorne (1977, p. 189):

> "The mutual independence of contemporaries constitutes their freedom. Without this independence, what happens anywhere would immediately condition what happens anywhere else. However, this would be fatal to freedom only if the sole alternative to mutual independence were mutual dependence. And this is not a necessary, it is even a possible, interpretation of Bell's result. What happens here now may condition what happens somewhere else without measurable temporal lapse, although what happens at somewhere else does not condition what happens here, still retains its freedom since ... no set of conditions can be fully determinative of the resulting actuality."

Quantum theory is formulated as an indeterministic theory. Each experimenter can choose freely which experiment he will perform. In addition, the result of the experiment is subject only to statistical requirements (Stapp, 2001b, p. 11):

> "These elements of 'freedom of choice', on the part of both the human participant and nature herself, lead to a picture of reality that gradually unfolds in response to choices that are not necessarily fixed by the prior physical part of reality alone. The central roles ... of these discrete choices ... make quantum theory a theory of discrete events, rather than a theory of the continuous evolution of locally conserved matter/energy."

The internal process of concrescence is not a spatiotemporal process. But the way in which the result of this internal process is "made available" to the external world is an atomic act. "Continuity is rejected as a basic feature of the units of becoming, but in the succession of the units of becoming what becomes is continuity" (McHenry, 2002, p. 168). Additionally, if quantum theory is a theory of observation, what does the term "observer" mean ? Physical instruments of measurement cannot be regarded as observers: They do not generate facts. One would come to a chain of observers. Where does this chain end ? Haag (2004, p. 55) comments:

> "Several eminent scientists (von Neumann, 1932; London and Bauer, 1939; Wigner, 1962) proposed that it terminates when an event becomes consciously perceived. Consciousness is regarded as the ultimate agency."

According to Heisenberg, "each occurring event signalizes a transition of the 'possible' to the 'actual' " (Stapp, 1979, p. 23). A becoming actual occasion receives past actual occasions as potentials for ingression into its own development. The development is one from potentiality to actuality and from actuality to potentiality. The potentials of past actual entities are interwoven into a unit by the activity of the growing actual entity. The newly grown actual entity is a real potential for future concrescence processes.

There are parallels between quantum theory and psychology. Stapp's "quantum theory of consciousness" is based on Heisenberg's interpretation that reality is a sequence of collapses of wave functions. Stapp observes that this view is similar to William James's view of mental life as "experienced sense objects". According to Stapp, the whole range of science, from atomic physics to mind-brain dynamics, is brought together in a single coherent theory of an evolving cosmos consisting of a physical reality with the closely related, but differently constituted, mental aspects of nature. Stapp holds that (Atmanspacher, 2006, p. 76)

> "Whiteheadian quantum ontology accepts ... the idea that our conscious intentions cause, at least in part, our intentional actions. This can be achieved by regarding the quantum reduction events to be the physical manifestations of the termination of psycho-physical process. ... The physical and psychological aspects of reality are thus tied together in the notion of a quantum event."

Is it now justified to argue that quantum events could be counted as sentient ? This assertion would equip elementary quantum events with a degree of creativity. It must first be asked how mentality is to be measured. One observes mentality concerning its effects out of the behavior of the things observed. To argue that each actual occasion possesses a mental pole is a consequence of the transference of human understanding to all events of nature. It conforms to the principle of unity of nature. Finally, quantum theory of consciousness as well as process philosophy delivers a rationally coherent way of understanding our conscious selves within the reality surrounding and sustaining us.

Whiteheadian quantum ontology is essentially an ontologization of the structure of orthodox relativistic quantum field theory, stripped of any anthropocentric formulations. (This means that mentality is no longer reserved for human beings and higher creatures.) But it is to a high degree anthropomorphic because this is the only way we can speak about reality. Thus, Whitehead's philosophy of organism is a logical transfer of the concepts of human experiences onto all entities of reality. In describing the last units of reality, he uses concepts which were derived from living organisms and applies them to the whole of nature.

Why is consciousness needed in the universe at all ? Because otherwise there would be no historical development. There were many possible changes from one state to another but no becoming anew. This leads to a "many-minds" picture "Each person's brain evolves quickly into [...] a smeared out continuum, and each stream of consciousness would be part of a continuous blur of classically describable possibilities" (Stapp, 2007, p. 59). The observed particularity would be the particularity of one individually observed branch of the universe. In this view, it is a property of each human consciousness to accommodate only a single one of these branches, even though all the branches exist together (Stapp, 1993, p. 188). The proposal of Heisenberg and Dirac as well as our human understanding assert the opposite: Nature actualizes one observable branch out of the emerging set of possible ones. The conflict originates from the continuous character of the description of nature provided by the quantum state and the discrete character of human experience. Real becoming necessitates temporal atomicity. But real temporality presupposes teleology, and consequently, mentality.

References

Atmanspacher, H. (2006): Clarifications and specifications. An interview with Henry Stapp. *Journal of Consciousness Studies* **13**(9), 67–85.

Burgers, J.M. (1965): *Experience and Conceptual Activity – A Philosophical Essay Based Upon the Writings of A.N. Whitehead*. MIT Press, Cambridge MA.

Cahill, R.T. (2005): Process physics: Self-referential information and experiential reality. Manuscript available at `www.ctr4process.org/publications/Articles/LSI05/Cahill-FinalPaper.pdf`.

Christian, W.A. (1959): *An Interpretation of Whitehead's Metaphysics*. Yale University Press, New Haven.

Eastman, E.E. and Keeton, H. (2004): *Physics and Whitehead – Quantum, Process, and Experience*. SUNY Press, New York.

Griffin, D.R. (1982): Parapsychology and Whitehead's philosophy. Audiotape recording available at `www.ctr4process.org/media/`.

Griffin, D.R. (2005): Whiteheadian physics: Its implications for time, consciousness, and freedom. Manuscript available at `www.ctr4process.org/publications/Articles/LSI05/Griffin-QuantumLecturefinal.pdf`.

Hättich, F. (2004): *Quantum Processes. A Whiteheadian Interpretation of Quantum Field Theory*. Agenda Verlag, Münster.

Haag, R. (2004): Quantum theory and the division of the world. *Mind and Matter* **2**(2), 53–66.

Hartshorne, C. (1977): Bell's theorem and Stapp's revised view of space-time. *Process Studies* **7**, 183–191.

Klose, J. (2002): *Die Struktur der Zeit in der Philosophie Alfred North Whiteheads*. Karl Alber, Freiburg.

London, F. and Bauer, E. (1939): *La théorie de l'observation en mécanique quantique*. Hermann, Paris.

Löw, R. (1980): *Philosophie des Lebendigen*. Suhrkamp, Frankfurt.

Lowe, V. (1941): William James and Whitehead's doctrine of prehension. *Journal of Philosophy* **38**, 113–126.

Malin, S. (2002): Whitehead's philosophy and quantum physics: A defense of *Nature Loves to Hide*. *Process Studies* **31.1**, 171–174.

Malin, S. (2006): What are quantum states? *Quantum Information Processing* **5**, 233–237.

McHenry, L. (2002): Whitehead, quantum mechanics and local realism. *Process Studies* **31.1**, 164–170.

Neumann, J. von (1932): *Mathematische Grundlagen der Quantenmechanik*. Springer, Berlin.

Rapp, F. (1986): Der Kreativitätsbegriff Whiteheads und die moderne Naturwissenschaft. In: Rapp, F. and Wiehl, R. (eds.), *Whiteheads Metaphysik der Kreativität*. Alber, Freiburg, pp. 81–104.

Schilpp, P.A. (1941): *The Philosophy of Alfred North Whitehead*. Open Court, La Salle IL.

Sherburne, D.W. (1966): *A Key to Whiteheads Process and Reality*. University of Chicago Press, Chicago.

Shimony, A. (1965): Quantum physics and the philosophy of Whitehead. In: Black, M. (ed.), *Philosophy in America*. Allen and Unwin, London, pp. 240–261. Republished in Shimony (1993b), pp. 291–309.

Shimony, A. (1993a): *Search for a Naturalistic Word View. Volume I. Scientific Method and Epistemology*. Cambridge University Press, Cambridge.

Shimony, A. (1993b): *Search for a Naturalistic Word View. Volume II. Natural Science and Metaphysics*. Cambridge University Press, Cambridge.

Stapp, H.P. (1977): Theory of reality. *Foundation of Physics* **7**, 313–323.

Stapp, H.P. (1979): Whiteheadian Approach to Quantum Theory and the Generalized Bell's Theorem. *Foundation of Physics* **9**, 1–25.

Stapp, H.P. (1993): *Mind, Matter, and Quantum Mechanics*. Springer, Berlin.

Stapp, H.P. (2001a): EPR-Bohr-Bell and nonlocality. Manuscript available at `www-physics.lbl.gov/~stapp/EPR.txt`.

Stapp, H.P. (2001b): Quantum theory and the role of mind in nature. *Foundations of Physics* **31**, 1465–1499.

Stapp, H.P. (2004): Whiteheadian process and quantum theory. In: Eastman, T.E. and Keeton, H. (eds.), *Physics and Whitehead – Quantum, Process, and Experience*. SUNY Press, New York, pp. 92–102.

Stapp, H.P. (2007): *Mindful Universe – Quantum Mechanics and the Participating Observer*. Springer, Berlin.

Whitehead, A.N. (1920): *The Concept of Nature*. Cambridge University Press, Cambridge.

Whitehead, A.N. (1925): *Science and the Modern World*. Macmillan, New York.

Whitehead, A.N. (1926): *Religion in the Making*. Macmillan, New York

Whitehead, A.N. (1929): *The Function of Reason*. Princeton University Press, Princeton.

Whitehead, A.N. (1933): *Adventures of Ideas*. Macmillan, New York.

Whitehead, A.N. (1938): *Modes of Thought*. Macmillan, New York.

Whitehead, A.N. (1978): *Process and Reality: An Essay in Cosmology*. Free Press, New York.

Wiehl, R. (1998): *Zeitwelten: philosophisches Denken an den Rändern von Natur und Geschichte*. Suhrkamp, Frankfurt.

Wiehl, R. (2000): Einleitung in die Philosophie A.N. Whiteheads. In: Wiehl, R. (ed.), *Abenteuer der Ideen*. Suhrkamp, Frankfurt, pp. 7–71.

Wigner, E.P. (1962): Remarks on the mind-body question. In: Good, I.J. (ed.), *The Scientist Speculates*. Heinemann, London, pp. 284–301.

Complementarity of Mind and Matter

Hans Primas

ETH Zürich, Switzerland, `primas@phys.chem.ethz.ch`

1 Introduction: Pauli on Mind and Matter

At the end of his authoritative article *The Influence of Archetypal Ideas on the Scientific Theories of Kepler* Wolfgang Pauli (1952, p. 164) stated:

> “The general problem of the relation between psyche and physis, between inside and outside, can hardly be regarded as solved by the term ‘psychophysical parallelism’ advanced in the last century.”

He continued with the visionary remark:

> “It would be most satisfactory of all if physis and psyche could be seen as complementary aspects of the same reality.”

Neither in Pauli’s publications nor in his letters we can find an elaboration of or more details about this idea. Consequently, I will not give an exegesis of Pauli’s writings, but use them just as a source of inspiration. First of all, we are confronted with the following four nontrivial question:

1. How can we characterize “physis”?
2. How can we characterize “psyche”?
3. What do we mean by “complementarity aspects”?
4. What do we mean by “the same reality”?

These questions are not yet well-defined – there are many possible answers. In the following I propose a scenario in which Pauli’s vision can be discussed consistently in simple and mathematically well-defined terms which allow us to work out additional details. Thereby I adopt an *ontological monism*, combined with an epistemic symmetry breaking leading to an *epistemic dual-aspect approach*. The dual aspects refer to tenseless descriptions in terms of a homogeneous time and to tensed descriptions in terms of a non-homogeneous time characterized by a privileged instant, the *Now*.

2 The First Problem: *What is Matter?*

2.1 Pauli's Answer

Even the answer to the apparently easy question "what is matter?" has changed dramatically several times since 1644, when Descartes classified matter and mind as distinct substances. At any rate, the traditional characterization of the physical as *res extensa* and of the mental as *res cogitans* does not allow to construct a workable theory for the mind–matter problem.

Similarly we cannot adopt atomism, which was the starting point of historical classical science: the idea that there exist theory-independent elementary objects, for example in the sense of Newton:[1]

> "God in the Beginning form'd Matter in solid,
> massy, hard, impenetrable moveable Particles".

Today nobody defends Newton's atomistic ontology any more. Nevertheless, the naive reductionism which tries to explain all phenomena in terms of entities at a supposedly lowest level of theoretical description is still popular. This approach fails simply *because the presumed lower level entities do not exist in a theory-independent sense.* Modern quantum mechanics put an end to atomism. The so-called "fundamental" entities (such as electrons, quarks, or gluons) represent *patterns of reality*, yet they are not building blocks of reality. They are not primary, but rather secondary and derived, in the same sense as solitons[2] are localized excitations of water, and not building blocks of water.

There is a bewildering diversity of concepts of matter.[3] Certainly no substantial advance in the mind–matter problem can be achieved without a clear characterization of what we mean by matter. Fortunately we can find a modern answer to the question *What is Matter?* in Pauli's (1954) contribution to the *International Symposium Presented in Honor of the Two-Hundredth Anniversary of Columbia University*:

> "Matter has always been and will always be one of the main objects of physics. ... even light has become matter now, due to Einstein's discoveries. It has mass and also weight; it is not different from ordinary matter, it too having both energy and momentum."
> "Taking the existence of all these transmutations into account, what remains of the old ideas of matter and substance? The answer is *energy*. This is the true substance, that which is conserved; only the form in which it appears is changing."

---

[1] Newton (1730), Quest. 31. In the Dover edition on p. 400.

[2] Solitons (discovered by Scott Russell in 1844) are localized steady two-dimensional waves of elevation that maintain their shape while propagating at the surface over a flat bed. They are important in our understanding of water waves.

[3] Compare for example the reviews by Weyl (1924) and by Johnson (1973).

2.2 Matter and Homogeneous Time

Adopting Pauli's answer, we have to recall that energy conservation is subject to a fundamental theorem by Emmy Noether (1918) which provides a deep connection between symmetries and conservation laws: *For every conservation law, there exists a continuous symmetry.* By definition, *energy* is the conserved quantity related to the *time-translation symmetry.* The concept of energy is meaningless if time is not homogeneous. In other words, *energy conservation holds only if the presupposed equations of motion do not contain any preferred moment of time.*

3 The Second Problem: *How to Characterize the Mental Domain?*

3.1 Violation of Physical Laws?

A recurrent theme in discussions of the mind–matter problem is the alleged violation of physical laws – usually without giving a precise account how we have to understand a "physical law".[4] Yet, the scope of validity of conservation laws is well understood: *a conservation law holds if and only if the system considered is invariant under the corresponding Noether-symmetry operation.* Therefore it makes no sense to argue as Bunge (1980, p. 17) does:

> "If immaterial mind could move matter, then it would create energy; and if matter were to act on immaterial mind, then energy would disappear. In either case energy ... would fail to be conserved. And so physics, chemistry, biology, and economics would collapse."

Intentional influences are not invariant under time-translations, so that in this case any argument involving energy conservation is misplaced. Whenever energy is well-defined, it is conserved by definition. Therefore it makes no sense to speak of the violation of energy conservation.

3.2 The Physical World is Not Causally Closed

Physicalism as a doctrine about the empirical world, claiming "that mental entities, properties, relations and facts are all really physical" (Crane and Mellor, 1990, p. 185). Or in another formulation: "Physics can, in principle, predict the probability with which a human body will follow any given trajectory" (Putnam, 1992, p. 83). Many philosophers assume a metaphysical naturalism,

---

[4] For example, Inwagen (1974, p. 188) admits that he does not know necessary and sufficient conditions for a proposition's being a law of nature or of physics, but generously he thinks that "we need not know how to analyze the concept 'law of physics'."

claiming that the realm of the physical is "causally closed".[5] In the weakest conception, causal closure means that "every physical phenomenon that has a sufficient cause has a sufficient physical cause" (Montero, 2003, p. 174). According to Kim (1980, p. 40) another way of stating the principle of physical causal closure is this:

> "If you pick any physical event and trace out its causal ancestry or posterity, that will never take you outside the physical domain. ... If you reject this principle, you are *ipso facto* rejecting the in-principle completability of physics."

This argument is not correct. It is true that under very general conditions any causally open physical system can be embedded into a causally closed time-invariant description with a larger state space.[6] However, such an *ad hoc* extension *presupposes a two-way determinism*, where the present is "mathematically determined jointly by the past and future, however remote" (Good, 1962). That is, only *if* the external influences *are given for all past and future times*, then we can reconstruct a local time-invariant deterministic description. Even if we can reconstruct a causal ancestry (which is by no means unique) in every particular case, this does not imply the possibility of a global causal reconstruction.

The assertion that "modern science is premised on the assumption that the material world is a causally closed system" (Heil, 1998, p. 23) is in striking contradiction to experimental science. Every experiment requires an irreversible dynamics. No experiment refers to a closed physical system. In a strictly deterministic world it would neither be possible to perform meaningful experiments nor to verify the partially causal behavior of a physical system. *We conclude that science neither assumes that the material world is a causally closed system, nor that physical laws imply the causal closure of physics.*

3.3 Experimental Physics Requires Intentionality

All experimental science is based on the understanding that the actions of an experimenter are *intentional*, and not actions which happen to him. There are no physical laws which cover intentionality (understood as the mind's directedness upon objects). *Experimental physics demands the distinction of past and future, the concept of the now, and the freedom of the experimenter to choose initial conditions.* To test *experimentally* whether a given physical system is causal, it is indispensable that the experimenter has the freedom *to deliberately choose* (within well-defined limits) a stimulus and then to record

---

[5] Without argument, Edelman (2004, p. 152) claims that "according to the laws of physics, the causal order is closed." However, as Hempel (1980) pointed out, the physical can neither be explained via reference to current physics, nor by reference to some future physics.

[6] Compare Howland (1974); Reed and Simon (1975), section X.12; Nickel (1996); Engel and Nagel (2000), section VI.9.

the response. Moreover, it is required that an experiment can be repeated at any particular instant.

Sometimes it is claimed that such a freedom is illusory. Yet, without this freedom all experimental science would be pointless:

To deny the freedom of action of an experimenter
is to deny the meaningfulness of experimental science.

Every experimental investigation presupposes that the specific design and implementation of an experiment is compatible with, *but not exclusively determined by, known physical laws.* This situation does not imply that the first principles of physics are inconsistent or not valid, but only that they cannot account for intentionally chosen experimental arrangements and initial conditions. This fact "proves that *contingency* is an essential feature of the world" (Weyl, 1952, p. 26).

3.4 A Mind–Matter Distinction

The empirically unfounded idea that the physical world is "causally closed" and that physical laws have universal validity is related to the failure to distinguish between *tenseless laws* and *tensed phenomena.* There are two classes of *conceptually different descriptions*:

- *Tensed descriptions* are characterized by a privileged position in time, called the *Now. Tensed time* is characterized by the Now and the indexical ordering of *past* and *future.*
- *Tenseless description* are characterized by a homogeneous time. *Tenseless time* is characterized by sequential orderings "earlier than" and "later than". Group-theoretically the homogeneity of tenseless time is represented by the invariance under time translations.

Tenseless time structures arise in theoretical physics. *In fundamental physics there is no flow of time.* All known fundamental principles of physics refer to laws which are invariant under time translations. Moreover, in the currently accepted fundamental laws of physics time-reversal symmetry is postulated to hold[7] so that for fundamental physical processes there is no preferred direction with respect to time. The homogeneity of tenseless time precludes the possibility of introducing a preferred instant into tenseless time, so that tenseless time cannot be reduced to tensed time, and conversely.

Since fundamental physics has no means to single out the notion of a Now, *the first principles of physics alone do not distinguish cause and effect*, and therefore cannot account for efficient causation.

The distinguishing quality of consciousness is the *Now* (the moment at which an experience takes place). Our experience of the *Now* does not reflect

---

[7] More precisely, this is true when time-reversal is combinded with parity-reversal and the interchange of particles and antiparticles.

a feature of some fundamental physical law. Therefore tenseless physics with a homogeneous time does not realize a complete theoretical scheme. What is missing is the concept of *Now*, necessary for the specification of initial conditions of physical experiments. For that reason it is rewarding to replace the Cartesian dualism of thinking and extended substances by an epistemic *duality of tensed and tenseless descriptions.*

This division of the world should be considered as a *regulative principle*, which is not given *a priori*. It is generated by an *epistemic* breaking of the holistic symmetry of the universe of discourse considered. Accordingly this distinction does not refer to nature itself. It is an organizing principle which is neither right nor wrong, but leads to merits and drawbacks of the chosen description. The division between tensed and tenseless descriptions is motivated by our desire to get context-independent and nonindexical universal "first principles". It corresponds to Newton's separation of *intentional initial conditions* from *universal laws of physics.*

4 The Third Problem: *Complementarity Beyond Physics?*

4.1 The Distinction Between A-Time and B-Time

Debates about the meaning of time go back to the ancients Greeks. The metaphysical dispute between Heraclitus and Parmenides continues to this day, with no decisive result.[8] McTaggart (1908) introduced a terminology that has been used ever since. The tensed indexical ordering of *past*, *now*, and *future* is referred to as *A-theoretical.* The tenseless pure sequential ordering *later than*, *simultaneous with*, and *earlier than* is referred to as *B-theoretical.* McTaggart held A-theoretical time to be more fundamental but inconsistent. Some philosophers claim that the "Now" is a real ontological feature of the world,[9] while some B-theorists claim that "past", "now" and "future" have no place in physics since they are only mind-dependent.[10] Furthermore, there is a great variety of different versions of A-time und B-time, and there are literally hundreds of most confusing philosophical papers discussing whether A-time or B-time adequately represents time, whether A-time is a special case of B-time or B-time is a special case of A-time, or whether A-time and B-time are indistinguishable. In addition there is an extensive but nonproductive philosophical rivalry between "presentism" (the doctrine that what does not

---

[8] For an anthology of the changing philosophical notions of time from the "Book of Genesis" to the work of twentieth-century philosophers compare Sherover (1975). For a recent survey of the controversies about an alleged "correctness" of A- or B-theories, compare Oaklander and Smith (1994).

[9] For example: Reichenbach (1956), Prior (1967), Gale (1968).

[10] For example: Williams (1951), Park (1972), Grünbaum (1973), Mellor (1998), Oaklander (2004).

exist in the present does not exist at all) and "eternalism" (the view that past, present, and future events are equally real).

4.2 Complementary Descriptions

In spite of the vast number of publications on this topic, I am not aware of any serious discussion of the obvious *possibility that the usual Boolean descriptions cannot encompass the totality of temporal phenomena.* This limitation of Boolean descriptions was already recognized more than hundred years ago by William James (1890, p. 206). He introduced the concept of *complementarity* to describe split modes of consciousness "which coexist but mutually ignore each other". Later, but still almost hundred years ago, Henri Bergson (1911, p. 344) distinguished two *complementary types of knowledge* of time: the one alludes to physical time, the other to our intuition of flowing time. In the 1930's, Niels Bohr (1948, p. 318) recognized that quantum mechanics offers

> "a novel relationship, conveniently termed complementarity, between empirical evidence obtained under different experimental conditions."

Reviewing the idea of complementarity in physics, Pauli (1948, p. 307) summarized:

> "Which knowledge is obtained and which other knowledge is irrevocably lost is left to the free choice of the experimentor, who may choose between mutually exclusive experimental arrangements."

Bohr never gave a formal definition of complementarity but restricted his discussion to the analysis of a number of examples. In the framework of traditional quantum mechanics the following characterization is often used: *two observables are complementary when the corresponding operators fail to commute.* More relevant than specific observables are the commutative algebras they generate. This fact suggests a slightly more general definition: *Two commutative subalgebras of an algebra of observables are said to be complementary if they cannot be embedded into a single commutative algebra.* Often it has been claimed that "complementarity distinguishes the world of quantum phenomena from the realm of classical physics".[11] That this view is conceptually misleading is evident by emphasizing the logical structure of complementarity. The following reformulation is equivalent for quantum physics, but applies far beyond quantum physics and includes the known examples from psychology, philosophy and engineering science:

Two Boolean descriptions are said to be complementary
if they cannot be embedded into a single Boolean description.

Here, a Boolean description refers to a Boolean domain where all propositions are *either true or false*, characterized by the *postulate of the excluded middle.* We know that a Boolean theory of matter is not possible, but we

---

[11] So begins a paper by Scully *et al.* (1991).

also know that Boolean descriptions play a privileged role since every experiment requires a *Boolean frame of reference*. As a consequence, a comprehensive description of the material domain requires a non-Boolean structure which is locally Boolean. The resulting non-Boolean structure is determined by the underlying partially overlapping Boolean algebras and the way they are pasted together. The most important example to get globally non-Boolean descriptions is to patch local Boolean descriptions together *smoothly*, in the same sense as geometric manifolds can be constructed out of Euclidean spaces (compare Primas, 2007).

In the following it will be shown that A-time descriptions and B-time descriptions can be embedded into a non-Boolean description such that they refer to well-defined *complementary* partially Boolean descriptions. To get a more precise characterization of A-time and B-time, we first discuss the possible symmetries of time structures.

5 The Fourth Problem: *How to Characterize the "Unus Mundus"?*

5.1 A Model of a Holistic Universe of Discourse

According to Carl Gustav Jung (1970, §767),

> "the idea of the *unus mundus* is founded on the assumption that the multiplicity of the empirical world rests on an underlying unity, and that not two or more fundamentally different worlds exist side by side or are mingled with one another. Rather, everything divided and different belongs to one and the same world, which is not the world of sense."

Jung borrowed the expression *unus mundus* from the Renaissance alchemist Gerhard Dorn. It is understood as an undivided primordial reality which is neutral with respect to the distinction of mind and matter, a "potential world outside time" (Jung, 1970, §718).

The *unus mundus* is perceived through Boolean reference frames but its full description is not embeddable into a Boolean framework. To comprehend physis and psyche "as complementary aspects of the same reality" we have to use a structure which supports locally Boolean descriptions but which is globally non-Boolean. There are many possible ways to do so. The best investigated structures of this type are the transitive partial Boolean algebras formed from the sets of all projections of non-commutative W*-algebras.[12]

---

[12] A W*-algebra is a *-algebra of bounded operators on a Hilbert space that is closed in the weak operator topology and contains the identity operator (W stands for "weakly closed").

5.2 The Nominal Distinction Between Mind and Matter

A globally non-Boolean description requires that the representing W*-algebra is a factor.[13] Consequently we describe our universe of discourse (considerend as a model for the *unus mundus*) by a factorial W*-algebra $\mathfrak{L}$.

For convenience but without loss of generality we divide the universe of discourse into three mutually compatible parts intended for the description of a material domain, a mental domain, and an interface between these two domains. The three mutually compatible parts of the universe of discourse will be represented by three kinematically independent W*-subalgebras $\mathfrak{M}$, $\mathfrak{T}$, and $\mathfrak{N}$ such that the W*-algebra $\mathfrak{L}$ of the universe of discourse is given by the following W*-tensor product[14]

$$\mathfrak{L} \;=\; \mathfrak{M} \mathbin{\bar{\otimes}} \mathfrak{T} \mathbin{\bar{\otimes}} \mathfrak{N}. \tag{1}$$

- The W*-algebra $\mathfrak{M}$ refers to a non-intentional *atemporal material domain.* In traditional quantum mechanics this algebra $\mathfrak{M}$ is taken as the algebra of all bounded operators acting on some Hilbert space, so that in this case $\mathfrak{M}$ is a factor of type I. In the following we adopt this choice.
- The W*-algebra $\mathfrak{N}$ refers to a *atemporal mental domain.* We consider *intentionality* as the mark of the mental. Even though we have no good information about how to specify the mental domain algebraically, we take it for granted that we can describe it in terms of the same fundamental logical structures as the atemporal material and the temporal domain.
- The W*-algebra $\mathfrak{T}$ refers to a *temporal domain*, defined as *interface* between the two atemporal domains. Thereby, concepts of time are supposed to describe the relation between the atemporal material and the atemporal mental domain. The structure of the temporal algebra $\mathfrak{T}$ will be discussed in the following section 6.

Since according to Einstein (1934, p. 165)

> "fundamental concepts and basic laws ... are free inventions of the human mind which admit no *a priori* justification either through the nature of the human mind or in any other way at all"

we consider the decomposition (1) as a regulative principle which has to be judged by its success. We adopt it on the ground that it is indispensable in physics.

---

[13] A W*-algebra is called a *factor* if its center contains only the scalar multiples of the identity. A factor is said to be of *type I*, if it contains an atom.

[14] If two algebras mutually commute element-wise they are said to be kinematically independent. If $\mathfrak{F}_1$ and $\mathfrak{F}_2$ are two kinematically independent factors of type I, then the factor generated by $\mathfrak{F}_1$ and $\mathfrak{F}_2$ is isomorphic to the W*-tensor product $\mathfrak{F}_1 \mathbin{\bar{\otimes}} \mathfrak{F}_2$.

6 Symmetries of the Temporal Domain

> "It has long been recognized that no analysis of natural science, whether it be physics or biology is complete unless we possess a proper analysis of its appropriate time-concept."
>
> NORBERT WIENER (1948, p. 197)

6.1 Preliminaries

There is no generally accepted answer to the old question "What is time itself?" *The reason is that "time" is a many-tiered concept that cannot be encompassed within a single Boolean description.*[15] Therefore time should not be regarded as a primitive notion but rather as a theoretical construct.

For our purpose the following idea by Leibniz provides a convenient *tentative* characterization of time:[16]

> "Given the existence of a multiplicity of concrete circumstances which are not mutually exclusive, we designate them as *contemporaneous* or *co-existing*. … *Time is the order of non-contemporaneous things*. It is thus the universal order of change in which we ignore the specific kind of changes that have occurred."

Although all fundamental notions of time are "free inventions of the human mind", they must be mapped onto the data given by empirical reality. In order that theoretical ideas about the nature of temporal phenomena do not degenerate into empty talk it is advisable to discuss first the mathematical time concepts which are well-established in engineering communication science and optics. Then we characterize the various time structures *group theoretically.*

First we have to bear in mind that we never measure time t but some indicators which we parametrize in terms of some clock variable $c(t)$ of an auxiliary real-valued parameter t, where the function $t \mapsto c(t)$ is chosen such "that the enunciation of the natural laws may be as simple as possible" (Poincaré, 1907, part I, chap. 2, sect. 13). Since natural laws neither distinguish an origin nor a scale of time, any reparametrization in terms of a clock variable $z \mapsto \tilde{c}(t) := c(at + \tau)$ with a real-valued parameter τ and $a > 0$ or $a < 0$ leads to an equivalent description of temporal phenomena. In this case a description of temporal phenomena in terms of an auxiliary parameter t is invariant under the group of linear transformations of the straight line,

$$t \;\rightarrow\; at + \tau, \quad \tau \in \mathbb{R}, \quad a > 0 \;\text{ or }\; a < 0. \tag{2}$$

---

[15] Even in natural science Rovelli (1995) distinguishes ten main notions of time and discusses their mathematical properties as used in physical theories.

[16] *Metaphysical Foundations of Mathematics*, 1715. Quoted from Wiener (1951), p. 202. The quoted minimal characterization does not yet imply Leibniz's somewhat problematic causal theory of time, which he expressed as follows: "When one of two non-contemporaneous elements contains the ground for the other, the former is regarded as the *antecedent*, and the latter as the *consequent*."

6.2 The Time–Frequency Phase Space

Since the real numbers $\mathbb{R}$ form an abelian group under addition,[17] the time-translations $t \to t+\tau$ generate a locally compact commutative one-parameter group $\mathscr{V} = \{\tau \in \mathbb{R}\}$, called the *time-translation group.* The set of all characters[18] on the time-translation group $\mathscr{V}$ is itself a locally compact commutative group, called the *dual group* $\widehat{\mathscr{V}}$ which is of crucial importance in communication theory. Every character of $\mathscr{V}$ has the form $\tau \mapsto e^{2\pi i\tau b}$ for $b \in \mathbb{R}$. For notational simplicity we denote the dual group $\widehat{\mathscr{V}}$ by $\mathscr{U} = \{b \in \mathbb{R}\}$.

As it is well known from engineering communication science and optics, signal analysis requires a self-dual two-dimensional phase space $\mathscr{V} \oplus \mathscr{U}$, where $\mathscr{V}$ is an additive locally compact commutative group, and $\mathscr{U} := \widehat{\mathscr{V}}$ is its dual. The time–frequency plane is parametrized by two complementary canonical coordinates $t \in \mathscr{V}$ (in engineering science called "time") and $\lambda \in \mathscr{U}$ (in engineering science called "frequency"). In the algebraic formulation the phase space $\mathscr{V} \oplus \mathscr{U}$ is replaced by a W*-algebra $\mathfrak{T}$ supporting an automorphic representation of the two groups $\mathscr{V}$ and $\mathscr{U}$.[19]

The phase space $\mathscr{V} \oplus \mathscr{U}$ admits only two types of integrable ergodic W*-systems[20] (Amann, 1986, p. 199), namely

(i) a Boolean system with the commutative eigenalgebra $\mathfrak{T}_{\mathrm{cl}} \simeq \boldsymbol{L}^\infty(\mathscr{V} \oplus \mathscr{U})$ generated by the commuting unitary operator groups $\{U_{\mathrm{cl}}(b)|b \in \mathscr{U}\}$ and $\{V_{\mathrm{cl}}(\tau)|\tau \in \mathscr{V}\}$,

$$U_{\mathrm{cl}}(b)\, V_{\mathrm{cl}}(\tau) \;=\; V_{\mathrm{cl}}(\tau)\, U_{\mathrm{cl}}(\tau), \quad b \in \mathscr{U}, \quad \tau \in \mathscr{V}, \tag{3}$$

(ii) a non-Boolean system with a factor $\mathfrak{T}$ of type I_∞, generated by the unitary operator groups $\{U(b)|b \in \mathscr{U}\}$ and $\{V(\tau)|\tau \in \mathscr{V}\}$ fulfilling the canonical commutation relation

$$U(b)\, V(\tau) \;=\; V(\tau)\, U(b)\, e^{2\pi i b\tau}, \quad b \in \mathscr{U}, \quad \tau \in \mathscr{V}. \tag{4}$$

---

[17] Recall that a group $(\mathscr{G}, \circ)$ is a set $\mathscr{G}$ on which an associative operation $\circ$ is defined, with a unit element e such that $e \circ g = g \circ e = g$ for all $g \in \mathscr{G}$, and such that every element g has an inverse.

[18] A character of a locally compact commutative group is an algebraic homomorphism of the group $\mathscr{G}$ into the multiplicative group of complex numbers with modulus 1. It can be realized as complex-valued continuous functions χ on $\mathscr{G}$ such that $\chi(g \circ g') = \chi(g)\, \chi(g')$ and $|\chi(g)| \;=\; 1$ $(g, g' \in \mathscr{G})$. The set of all characters on $\mathscr{G}$ is itself a locally compact commutative group, called the dual group $\widehat{\mathscr{G}}$ of $\mathscr{G}$.

[19] An automorphisms of a *-algebra $\mathfrak{K}$ is a structure preserving map, i.e. a *-isomorphism of $\mathfrak{K}$ onto itself. An automorphic action α of a group $\mathcal{G}$ on a *-algebra $\mathfrak{K}$ is a homomorphism $g \mapsto \alpha_g$, such that each α_g is an automorphism of $\mathfrak{K}$, so that $\alpha_{gg'} = \alpha_g \circ \alpha_{g'}$, $\alpha_g(XY) = \alpha_g(X)\, \alpha_g(X)$, and $\alpha_g(X^*) = \alpha_g(X)^*$.

[20] A W*-system $(\mathfrak{T}, \mathscr{G}, \alpha)$, where $\mathfrak{T}$ is a W*-algebra, $\mathscr{G}$ a group and α an automorphisms, is said to be *ergodic* if $\alpha_g(X) = X$ for all $g \in \mathscr{G}$ and $X \in \mathcal{T}$ implies that X is a multiple of the identity operator.

A theory of material phenomena in terms of a *parameter time* requires a description of the type (i) with a *Boolean* description of time phenomena. In this case the selfadjoint generator of the unitary operator $U_{\mathrm{cl}}(b) = \exp(2\pi i b T_{\mathrm{cl}})$ defines a *classical time operator* T_{cl} whose values correspond to the parameter time.

Descriptions in terms of a classical time operator cannot describe all aspects of time, like quantum interferences in time.[21] A fully holistic description of matter requires the non-Boolean description (ii) of the temporal domain, where the algebras $\mathfrak{T}$ and $\mathfrak{M}$, hence also $\mathfrak{T} \bar{\otimes} \mathfrak{M}$, are factors. In this description the unitary operators $V(\tau) \in \mathfrak{T}$ and $U(b) \in \mathfrak{T}$ do not commute, so that the selfadjoint time operator T defined by

$$U(b) \;=\; e^{2\pi i b T}, \quad b \in \mathscr{U}\,. \tag{5}$$

cannot have a sharp value. The extended time operator $\mathbf{1} \otimes T$ commutes with all extended observables $M \otimes \mathbf{1}$ $(M \in \mathfrak{M})$ of the material system, so that it is *not* canonically conjugate to the Hamiltonian of the material system.

6.3 Symmetries Related to Time Phenomena

Symmetries serve as guides in the construction of theories. Symmetry is here understood as invariance under a group of transformations. Modern signal theory is based on the group-theoretical ideas behind harmonic analysis, an important tool in communication engineering and information theory.[22] Within the temporal domain the so-called *affine Weyl–Heisenberg group* plays a crucial role. It can be described in terms of the following three commutative symmetry groups:

- The *time-translation group* $\mathscr{V}$ describes time shifts $t \to t + \tau$. On the algebra $\mathfrak{T}$ it is realized by the unitary one-parameter group $\{V(\tau) | \tau \in \mathscr{V}\}$. This group is isomorphic to the additive group $\mathbb{R}$ of real numbers, and fulfills the group relation $V(\tau)V(\tau') = V(\tau+\tau')$. The selfadjoint generator Λ of the unitary operator $V(\tau) = e^{-2\pi i \tau \Lambda}$ is called the *frequency operator*. Invariance under time-translations implies conservation of energy.
- The *frequency-translation group* $\mathscr{U}$ describes the frequency shifts $\lambda \to \lambda + b$. On the algebra $\mathfrak{T}$ it is realized by the unitary one-parameter group $\{U(b) | b \in \mathscr{U}\}$. This group is isomorphic to the additive group $\mathbb{R}$ of real numbers, and fulfills the group relation $U(b)U(b') = U(b+b')$. The selfadjoint generator T of the unitary operator $U(b) = e^{2\pi i b T}$ is called the *time operator*, which is responsible for the ordering of temporal phenomena.

---

[21] Novel versions of the double-slit interference experiment with slits separated in time prove the existence of quantum interferences in time. Compare Szriftgiser *et al.* (1996), Wollenhaupt *et al.* (2002), Lindner *et al.* (2005). Such experiments cannot be explained in terms of a classical time operator.

[22] Compare for example Cohen (1995), Hogan and Lakey (2005).

- The *scaling group* $\mathscr{X}$ describes the scaling operation $t \to at\,, \lambda \to \lambda/a$ for $a > 0$. On the algebra $\mathfrak{T}$ it is realized by the unitary one-parameter group $\{X(a)|a > 0\}$. This group is isomorphic to the multiplicative group $\mathbb{R}^+$ of positive numbers and fulfills the group relation $X(a)X(a') = X(aa')$. The selfadjoint generator S of the unitary operator $X(a) = e^{2\pi i \ln(a) S}$ is called the *scale operator*. Scaling invariance means that there is no universal unit of time or frequency.

Why is the scaling group important? Invariance under scale transformation plays a crucial role for the recognition of temporal patterns. An outstanding fact of our mental structure is the ability to abstract from irrelevant attributes of a signal and be aware only of *invariant features*. For example, our ability to recognize the same piece of music played at different time scales implies that auditory perception is invariant under moderate scalings. A composer does neither deal with the time representation nor with the frequency representation, but he produces a representation in the form of a musical score which is manifestly scale invariant.

The well-known mathematical fact that a signal which is confined to a small time interval cannot be confined to a small frequency interval is of great engineering importance.[23] In particular, there are no signals which are both limited in time and frequency. However, an *approximately* band-limited function can be *approximately* time-limited. In 1928 Ralph Vinton Lyon Hartley (1928, p. 535) showed that "the amount of information which can be transmitted is proportional to the product of the width of the frequency range by the time it is available."

Much later Slepian, Landau and Pollak[24] were able to rephrase the somewhat vague formulation by Hartley by a rigorous mathematical theorem saying that the number of orthogonal signals of approximative bandwidth $\Delta\lambda$ and approximative time duration $\Delta\tau$ is approximately equal $2(\Delta\lambda)(\Delta\tau)$. If we define a non-probabilistic Hartley information by $H = \log\{2(\Delta\lambda)(\Delta\tau)\}$, then this result implies that the amount of information of a signal is *invariant under scaling transformations*. That is, for every $a > 0$ the signals $t \mapsto s(t)$ and $t \mapsto \sqrt{a}\, s(at)$ have the same Hartley-information capacity. To summarize: *Hartley information is the conserved quantity associated with scale invariance.*

6.4 The Affine Weyl–Heisenberg Group

The fact that we cannot independently *control* the time behavior and the frequency behavior of a signal shows that the time description and the frequency

---

[23] The fact that a nonzero function with compact support cannot have a Fourier transform with compact support can be formulated in many different mathematically rigorous ways, compare for example the monograph by Havin and Jöricke (1994).

[24] Slepian and Pollak (1961), Landau and Pollak (1961), Landau and Pollak (1962). Compare also Slepian (1976), Slepian (1983).

description cannot be embedded into a single Boolean description. The commutative time-translation group $\mathscr{V}$, the commutative frequency-translation group $\mathscr{U}$, and the commutative scaling group $\mathscr{X}$ are the bases of the various time–frequency–scale representations commonly used for signal processing and wavelet analysis. Each of these commutative groups formalizes a *Boolean feature* of time phenomena. Studying a signal only from one of these aspects cannot reveal all the information a signal carries. The three commutative one-parameter groups $\mathscr{U}$, $\mathscr{V}$, and $\mathscr{X}$ represent three mutually complementary features in the sense that they cannot be embedded into a single Boolean description.

This situation requires a generalization of the harmonic analysis of commutative groups. The non-Boolean structure of time with its Boolean aspects described by the three commutative groups $\mathscr{U}$, $\mathscr{V}$, and $\mathscr{X}$ is encoded in the so-called *affine Weyl–Heisenberg group*.[25] To construct this group from the groups $\mathscr{U}$, $\mathscr{V}$ and $\mathscr{X}$ we first observe that according to eq.(4) the unitary operators $U(b)$ and $V(\tau)$ are not closed under multiplications, so that it is convenient to introduce a unitary operator $W_0(b,\tau,\varphi)$ by

$$W_0(b,\tau,\varphi) \;:=\; e^{2\pi i\varphi}\, e^{\pi i b\tau}\, V(\tau)\, U(b), \tag{6}$$

where $\varphi \in \mathbb{R}$ is an auxiliary toral component. These operators fulfill the commutation relation

$$W_0(b,\tau,\varphi)\, W_0(b',\tau',\varphi') \;=\; W_0\big(b+b',\tau+\tau',\varphi+\varphi'+\tfrac{1}{2}(b\tau'-b'\tau)\big),$$

so that $\{W_0(b,\tau,\varphi) | b \in \mathscr{U}, \tau \in \mathscr{V}, \varphi \in \mathbb{R}\}$ represents a non-commutative group, called the *Weyl–Heisenberg group* $\mathscr{W}_0$.[26]

The *affine Weyl–Heisenberg group* $\mathscr{W}$ can be viewed as the extension of the Weyl–Heisenberg group $\mathscr{W}_0$ incorporating dilations. It is a unimodular non-commutative group with the group law

$$\begin{aligned} &W(a,b,\tau,\varphi)\, W(a',b',\tau',\varphi') \;= \\ &W(aa', b+ab', \tau+\tau'/a, \varphi+\varphi'+\tfrac{1}{2}[b\tau'/a - ab'\tau]). \end{aligned} \tag{7}$$

---

[25] The harmonic analysis of the non-commutative Weyl–Heisenberg group is a well-established tool to discuss temporal phenomena in signal theory, compare for example Howe (1980), Schempp (1986), Gröchenig (2001). The extension of the Weyl-Heisenberg group by dilations is called the affine Weyl–Heisenberg group; compare Grossmann *et al.* (1986), Torrésani (1991), Kalisa and Torrésani (1991). For applications of covariant time–frequency–scale analysis compare for example Hlawatsch *et al.* (2003) and Papandreou-Suppappola (2003).

[26] The Weyl–Heisenberg group is a simply connected and nilpotent unimodular three-dimensional Lie group with the underlying manifold $\mathbb{R}^3$, whose Haar measure coincides with the Lebesgue measure on $\mathbb{R}^3$. The set $\{W_0(0,0,\varphi) | \varphi \in \mathbb{R}\}$ is a central closed subgroup $\mathscr{Z}_0$ of $\mathscr{W}_0$. The quotient group $\mathscr{W}_0/\mathscr{Z}_0$ is called the *reduced Weyl–Heisenberg group*. It can be identified with the group $\mathbb{R}^2$ with the Haar measure $db\, d\tau$.

The unitary operators $W(a,b,\tau,\varphi)$ can be represented by a suitable combination of the unitary scaling operator $X(a)$, the unitary frequency-shift operator $U(b)$, and the unitary time-translation operator $V(\tau)$:

$$W(a,b,\tau,\varphi) \;=\; e^{2\pi i\varphi}\, e^{\pi i b\tau}\, V(\tau)\, U(b)\, X(a). \tag{8}$$

Some relations for the affine Weyl–Heisenberg group

Besides the trivial relations $X(a)X(a') = X(aa')$, $U(b)U(b') = U(b{+}b')$ and $V(\tau)V(\tau') = V(\tau + \tau')$, the unitary operators X, U, V fulfill the following commutation relations

$$\begin{aligned} X(a)\,U(b) &= U(ab)\,X(a) \quad , && \text{(9a)}\\ V(\tau)\,X(a) &= X(a)\,V(a\tau) \quad , && \text{(9b)}\\ U(b)\,V(\tau) &= e^{2\pi i b\tau} V(\tau)\,U(b) \quad . && \text{(9c)} \end{aligned}$$

The commutation relations (9) imply the following transformation relations:

$$\begin{aligned} e^{-2\pi i \ln(a)S}\, T\, e^{2\pi i \ln(a)S} &= a^{-1}\,T \quad , \quad a > 0 \quad , && \text{(10a)}\\ e^{-2\pi i \ln(a)S}\, \Lambda\, e^{2\pi i \ln(a)S} &= a\,\Lambda \quad , \quad a > 0 \quad , && \text{(10b)}\\ e^{2\pi i \tau\Lambda}\, S\, e^{-2\pi i \tau\Lambda} &= S + \tau\,\Lambda \quad , \quad \tau \in \mathbb{R} \quad , && \text{(10c)}\\ e^{2\pi i \tau\Lambda}\, T\, e^{-2\pi i \tau\Lambda} &= T + \tau\,\mathbf{1} \quad , \quad \tau \in \mathbb{R} \quad , && \text{(10d)}\\ e^{-2\pi i bT}\, S\, e^{2\pi i bT} &= S + b\,T \quad , \quad b \in \mathbb{R} \quad , && \text{(10e)}\\ e^{-2\pi i bT}\, \Lambda\, e^{2\pi i bT} &= \Lambda + b\,\mathbf{1} \quad , \quad b \in \mathbb{R} \quad . && \text{(10f)} \end{aligned}$$

7 Contextually Broken Symmetries

> “Zweiteilung und Symmetrieverminderung,
> das ist des Pudels Kern.”
> Wolfgang Pauli[27]

7.1 On the Necessity of Breaking the Holistic Symmetry

In a full-fledged non-Boolean universe of discourse there are no patterns. Any knowledge of the world depends on a particular choice of perspective. Furthermore, all well-defined evidence “must be expressed in ordinary language making use of common logic” (Bohr, 1948, p. 317) so that any description requires a Boolean frame of reference.[28] The selection of such a Boolean frame of reference implies that we have to decide which features we consider as relevant. Unless we reduce the holistic symmetry by suppressing “irrelevant” features we do not get recognizable patterns.

---

[27] In a letter to Heisenberg. Quoted from Heisenberg (1959), p.663.

[28] For a discussion in terms of partial Boolean algebras compare Primas (2007).

What we mean by the "holistic symmetry of the universe of discourse" can only be understood in terms of symmetry breakings. In a perfectly symmetric situation there are no distinctions, so that reality does not appear in a structured form. Fundamental symmetries are never directly accessible, they can only be retrospectively inferred by contextual symmetry breakings. The conceptual necessity to break symmetries was clearly recognized long ago by Pierre Curie (1894):

> "C'est la dissymétrie qui crée le phénomène."

The main enigma of any description of a patternless *unus mundus* is to find appropriate partitions which create relevant patterns. By reducing the symmetry of the universe of discourse we partition it into domains which are amenable to an epistemic description. Clearly such an epistemic symmetry breaking implies hat every description of a holistic universe can be a partial description at most.

7.2 Subgroups of the Affine Weyl–Heisenberg Group

The most important symmetries of the temporal domain are covered by the affine Weyl–Heisenberg group $\mathscr{T}$. In this holistic representation directly ascertainable temporal phenomena are not manifest. To make them manifest, *we have to reduce the full symmetry of the affine Weyl–Heisenberg group* to one of its subgroups. One way to study contextually broken symmetries described by a subgroup $\mathscr{G}$ of the full symmetry group $\mathscr{T}$ of the temporal domain is to decompose the representation of $\mathscr{G}$ into elementary building blocks.

If we select a particular subgroup $\mathscr{G} \subset \mathscr{T}$, a continuous temporal symmetry is represented by a mapping $\alpha : \mathscr{G} \to \mathrm{Aut}(\mathfrak{T})$ of the symmetry group $\mathscr{G}$ into the automorphism of the algebra $\mathfrak{T}$ of the temporal domain. The W*-algebra

$$\mathfrak{T}_\alpha \;:=\; \big\{ F \in \mathfrak{T} \,\big|\, \alpha_g(F) = F \text{ for all } g \in \mathscr{G} \big\} \tag{11}$$

is called the *fixed point algebra* of the W*-system $(\mathfrak{T}, \alpha, \mathscr{G})$. If the fixed point algebra is trivial, $\mathfrak{T}_\alpha = \mathbb{C}$, the W*-system $(\mathfrak{T}, \alpha, \mathscr{G})$ is group-theoretically *indecomposable* or, mathematically speaking, *ergodic*. Ergodic systems are conceptually of crucial importance, they represent *elementary systems* which determine in a mathematically and conceptually sound manner the relevant observables of the description associated with the group $\mathscr{G}$ (Amann, 1986). Moreover, if $\mathscr{G}$ acts ergodically on an algebra, this algebra is necessarily a *factor* (Kadison and Ringrose, 1986, pp. 546–547). If we can decompose the representation $\alpha : \mathscr{G} \to \mathrm{Aut}(\mathfrak{T})$ into ergodic ones, we obtain the central decomposition of the W*-algebra $\mathfrak{T}$ into a direct sum or a direct integral of factors, that is into W*-algebras with trivial centers. Since the center $\mathfrak{Z}(\mathfrak{T})$ of $\mathfrak{T}$ is a *commutative* algebra,

$$\mathfrak{Z}(\mathfrak{T}) \;:=\; \{ Z \,|\, Z \in \mathfrak{T} \text{ with } ZF - FZ \text{ for every } F \in \mathfrak{T} \}, \tag{12}$$

the elements $\mathfrak{Z}(\mathfrak{T})$ lead to a *Boolean classification of temporal phenomena.*

In the following two sections we will illustrate the emergence of specific time structures by symmetry reduction leading to the two complementary affine subgroups $\mathscr{A}$ and $\mathscr{B}$,

$$\mathscr{A} := \{A(a,b) | a > 0, b \in \mathscr{U}\} \text{ with } A(a,b) = W(a,b,0,0)\,, \tag{13a}$$

$$\mathscr{B} := \{B(a,\tau) | a > 0, \tau \in \mathscr{V}\} \text{ with } B(a,\tau) = W(a,0,\tau,0), \tag{13b}$$

of the affine Weyl–Heisenberg group $\mathscr{W}$. We use them to represent *complementary features of the time-structure of the temporal domain.*

8 Descriptions in Terms of A-Time

> "The experience of the Now means something special for man, something essentially different from the past and the future, but this important difference does not and cannot occur within physics."
>
> ALBERT EINSTEIN (quoted in Carnap, 1963, p. 37)

8.1 Time *in Statu Nascendi*

A-time always refers to processes of becoming. This openness and the directionality of the dynamic A-time can be captured by Weyl's *medium of free becoming*,[29] where "*time* as the most fundamental continuum" (Weyl, 1918, p. 67) is taken in the intuitionist interpretation of Brouwer and Weyl. This continuum is "*created step by step by free acts of choice*, and thus necessarily remains *in statu nascendi*" (Weyl, 1949, p. 52). According to Bernays (1935, p. 63) "the characteristic general feature of intuitionism for the whole development of science does not result from pure intuition, but rather from the understanding of the reflecting and acting subject".

8.2 Elementary A-Time Systems

The affine group $\mathscr{A}$ defined by (13a) fulfills the group relation

$$A(a,b)\,A(a',b') = A(aa', b + ab'). \tag{14}$$

It describes the *tensed A-time* and allows a precise description of the concepts "past" and "future". On the non-commutative algebra $\mathfrak{T}$ (generated by the two non-commuting unitary operator groups $\{U(b)|b \in \mathscr{U}\}$ and $\{V(\tau)|\tau \in \mathscr{V}\}$) the affine group $\mathscr{A}$ is represented by the unitary operators

---

[29] Weyl (1949), p. 52: "The continuum no longer appears, to use Leibniz's language, as an aggregate of fixed elements but as a medium of free 'becoming'." For more details, compare Weyl (1925).

$$A(a,b) \;=\; U(b)\,X(a), \quad a>0, \quad b\in\mathscr{U}, \tag{15}$$

where $\{U(b)|b\in\mathscr{U}\}$ and $\{X(a)|a>0\}$ represent the frequency-translation and the scaling group, respectively. Since $\mathfrak{T}$ is a factor, the automorphisms generated by the affine group $\mathscr{A}$ are inner, so that they can be implemented by the unitary operators $A(a,b)$. The fixed point algebra of the action of $\mathscr{A}$ on $\mathfrak{T}$ is given by

$$\mathfrak{A} \;:=\; \big\{F\in\mathfrak{T}\,\big|\,A(a,b)FA(a,b)^* = F \text{ for all } a>0\,, b\in\mathscr{U}\big\}. \tag{16}$$

The well-known representation theory of affine groups[30] implies that the W*-algebra $\mathfrak{A}$ generated by the affine group $\{A(a,b)|a>0, b\in\mathscr{U}\}$ is the direct sum of three factors

$$\mathfrak{A} \;=\; \mathfrak{A}_-\oplus\mathfrak{A}_0\oplus\mathfrak{A}_+, \tag{17}$$

where $\mathfrak{A}_0$ is the trivial W*-algebra of complex numbers $\mathbb{C}$. On the infinite-dimensional algebra $\mathfrak{A}_\pm$ the affine group $\mathscr{A}$ acts ergodically. The center of $\mathfrak{A}$ is generated by three orthogonal projectors P_-, P_0 and P_+, defined by

$$P_0\,\mathfrak{A}\,P_0 \;=\; \mathfrak{A}_0, \quad P_\pm\,\mathfrak{A}\,P_\pm \;=\; \mathfrak{A}_\pm. \tag{18}$$

The projectors P_-, P_0 and P_+ commute with all observables in $\mathfrak{A}$ and have dispersion-free values in each pure state of the A-time system. For this reason they are referred to as *classical observables* with respect to the A-time description.

Since the affine group $\mathscr{A}$ acts ergodically on the factors $\mathfrak{A}_0, \mathfrak{A}_\pm$, the time operator T and the scale operator S an be decomposed as

$$T = T_+\oplus T_0\oplus T_-, \quad S = S_+\oplus S_0\oplus S_-, \tag{19}$$

with the restrictions

$$T_0 \;=\; T|_{\mathfrak{A}_0}, \quad T_\pm \;=\; T|_{\mathfrak{A}_\pm}, \quad S_0 \;=\; S|_{\mathfrak{A}_0}, \quad S_\pm \;=\; S|_{\mathfrak{A}_\pm}. \tag{20}$$

The operator T_+ is positive, while T_- is negative, so that the time concept related to the affine group $\mathscr{A}$ allows a precise description of the concepts "past", "punctual Now" and "future" of A-time:

the time operator $T_- < 0$ characterizes the *past*, (21a)

the time operator $T_0 = 0$ characterizes the *punctual Now*, (21b)

the time operator $T_+ > 0$ characterizes the *future.* (21c)

Since the classical observables P_+, P_0 and P_- generate the center of $\mathfrak{A}$, the factors $\mathfrak{A}_-$, $\mathfrak{A}_o$ and $\mathfrak{A}_-$ are separated by a superselection rule which prohibits coherent superpositions between the past, the punctual Now and the future. That is, *a description in terms of A-time corresponds to the traditional distinction between past and future, and does not allow the possibility of holistic correlations between past and future.*

---

[30] Gelfand and Neumark (1947). Compare also Vilenkin (1968), chapter V, §1.

8.3 Breakdown of Translation Symmetry and Reversal Symmetry

On the algebra $\mathfrak{T}$ time is represented by the one-parameter group $\{\theta_\tau | \tau \in \mathbb{R}\}$ of automorphisms θ_τ, defined by

$$\theta_\tau(Y) \;:=\; V(\tau)^* \, Y \, V(\tau), \quad Y \in \mathfrak{T}, \quad \tau \in \mathbb{R}. \tag{22}$$

where $\{V(\tau) | \tau \in \mathscr{V}\}$ is the unitary *time-translation group* $\mathscr{V}$. The restriction of the automorphisms θ_τ to the subalgebra $\mathfrak{A}_\pm$ is no longer an automorphism but an endomorphism[31] $\theta_\tau^\pm$,

$$\theta_0^\pm \;=\; id, \quad \theta_\sigma^\pm \circ \theta_\tau^\pm \;=\; \theta_{\sigma+\tau}^\pm, \quad \sigma, \tau \geq 0, \tag{23}$$

which is implemented by the isometric semigroup[32] $\{V_\pm(\tau) | \tau \geq 0\}$,

$$\theta_\tau^\pm(A_\pm) \;=\; V_\pm^*(\tau) \, A_\pm \, V_\pm(\tau), \quad A_\pm \in \mathfrak{A}_\pm, \quad \tau \geq 0, \tag{24}$$

where $V_\pm(\tau)$ is the restriction of the unitary operator $V(\tau)$ to the subalgebra $\mathfrak{A}_\pm$,

$$V_\pm(\tau) \;:=\; V(\pm\tau)|_{\mathfrak{A}_\pm}, \quad \tau \geq 0. \tag{25}$$

With respect to the elementary A-time systems, *the time-translation symmetry is broken*, but the corresponding factors $\mathfrak{A}_+$ and $\mathfrak{A}_-$ are still invariant under the action of the *semigroup* $\{\theta_\tau^\pm | \tau \geq 0\}$

$$\theta_\tau^\pm(\mathfrak{A}_\pm) \;\subseteq\; \mathfrak{A}_\pm \quad \text{for} \quad \tau \geq 0. \tag{26}$$

Moreover, the A-time operators T_+ and T_- still transform covariantly under endomorphism $\theta_\tau(A_\pm)$:

$$\theta_\tau(T_\pm) \;=\; T_\pm \,\pm\, \tau \, \mathbf{1}, \quad \tau \geq 0. \tag{27}$$

Since the semigroup $\{V_+(\tau)\}$ is strongly contractive, $V_+(\tau) \to 0$ as $\tau \to \infty$, it describes the directed flow of time of a *decaying* system. For this reason $\mathfrak{A}_+$ is called the *outgoing subalgebra*, and $\mathfrak{A}_-$ is called the *incoming subalgebra*.

In the ergodic A-time description not only the time-translation symmetry but also *the time-reversal symmetry is broken*. The operation $t \to -t$ swaps the subalgebras $\mathfrak{A}_+$ and $\mathfrak{A}_-$, and the two semigroups $\{V_+(\tau) \,|\, \tau \geq 0\}$ and $\{V_-(\tau) \,|\, \tau \geq 0\}$ are related by time inversion,

$$V_+(\tau) \;:=\; P_+ V(\tau) P_+, \quad V_-(\tau) \;:=\; P_- V(-\tau) P_-, \quad \tau \geq 0. \tag{28}$$

---

[31] An endomorphism of a W*-algebra $\mathfrak{X}$ is a linear map $\gamma : \mathfrak{X} \to \mathfrak{X}$ with $\gamma(XY) = \gamma(X) \circ \gamma(Y)$ for all $X, Y \in \mathfrak{X}$, which preserves the *-operation, $\alpha(X^*) = \alpha(X)^*$ for all $X \in \mathfrak{X}$.

[32] More precisely, $\{V_\pm(\tau) \,|\, \tau \in \mathscr{V}\}$ is a completely nonunitary and strongly continuous semigroup of contractive isometries, so that $V_\pm(\tau)V_\pm(\tau') = V_\pm(\tau + \tau')$ for $\tau, \tau' \geq 0$ and $\|V_\pm(\tau)\| \leq 1$ for all $\tau \geq 0$. A contraction semigroup $\{V_\pm(\tau)\}$ is said to be completely nonunitary if there exists no nontrivial subspace reducing all $V_\pm(\tau)$ in which $V_\pm(\tau)$ acts unitarily.

8.4 Canonical Operators for the A-Time Description

The commutation relation (9a) implies the *affine commutation relation*

$$S\,T \;-\; T\,S \;=\; (1/2\pi i)\,T \tag{29}$$

for the time operator T and the scale operator S. The general inequality $\Delta X \Delta Y \geq \frac{1}{2}\big|XY - YX\big|$ implies the following affine uncertainty relation

$$\Delta T \Delta S \;\geq\; \frac{1}{4\pi}\,\big|\langle T\rangle\big|, \tag{30}$$

where $\langle\cdots\rangle$ denotes the expectation value, $(\Delta T)^2 := \langle T^2\rangle - \langle T\rangle^2$, and $(\Delta S)^2 := \langle S^2\rangle - \langle S\rangle^2$.

By restricting the affine commutation relations to the algebras $\mathfrak{A}_+$ and $\mathfrak{A}_-$ we obtain

$$S_\pm\,T_\pm \;-\; T_\pm\,S_\pm \;=\; (1/2\pi i)\,T_\pm. \tag{31}$$

On the invariant subalgebras $\mathfrak{A}_+$ and $\mathfrak{A}_-$ one can define selfadjoint operators fulfilling the usual *canonical* commutation relations. Using the generalized function[33] $x \mapsto (\ln|x|)_\pm$, the *logarithmic time operator* $(\ln T)_\pm$ and the scale operator $S_\pm$ fulfill the canonical commutation relations

$$S_\pm\,(\ln T)_\pm \;-\; (\ln T)_\pm\,S_\pm \;=\; (1/2\pi i)\,\mathbf{1}. \tag{32}$$

It is straightforward to prove that the operator pair $\{T, \Lambda\}$ and the operator pair $\{\ln|T|_\pm\,, S_\pm\}$ are unitarily equivalent. While the canonical pair $\{T, \Lambda\}$ generates the W*-algebra $\mathfrak{T}$ of the temporal domain, the canonical pair $\{\ln|T|_\pm\,, S_\pm\}$ generates the W*-algebra $\mathfrak{A}_\pm$ for the elementary A-time-description. The canonical commutation relation (32) implies the following inequality for the variances of the logarithmic time and the scale operator,[34]

$$\Delta(\ln T)_\pm\;\Delta S_\pm \;\geq\; (1/4\pi). \tag{33}$$

Therefore neither $(\ln T)_\pm$ nor $S_\pm$ can have sharp values. Nonetheless, in the A-time-description the past and the future are sharply separated by a superselection rule.

---

[33] $(\ln|x|)_\pm := \ln(|x|)$ for $\pm\,x > 0$ and $(\ln|x|)_\pm := 0$ for $\pm\,x < 0$. Compare Kanwal, 1983, pp. 86–88.

[34] $(\Delta(\ln T)_\pm)^2 \;:=\; \langle(\ln T)^2_\pm\rangle - \langle(\ln T)_\pm\rangle^2$, $(\Delta S_\pm)^2 \;:=\; \langle S^2_\pm\rangle - \langle S_\pm\rangle^2$, where the mean values of the operators $(\ln T)_\pm$, $(\ln T)^2_\pm$, $S_\pm$, $S^2_\pm$ are denoted by $\langle(\ln T)_\pm\rangle$, $\langle(\ln T)^2_\pm\rangle$, $\langle S_\pm\rangle$, $\langle S^2_\pm\rangle$, respectively.

9 Descriptions in Terms of B-Time

> "The objective world simply *is*, it does not *happen*."
> HERMANN WEYL (1949, p. 116)

9.1 The Rigid Structure of B-Time

A-time and B-time are based on conceptually complementary views on the nature of the real numbers: the intuitionistic–constructivistic point of view (exemplified by the geometric concept of a straight line), and Cantor's platonistic concept of the atomistic continuum (exemplified by the concept of whole numbers).[35] According to Paul Bernays the intuition of continuity and the intuition of discreteness refer to two different kinds of existence which require complementary perspectives: "the objective, theoretical (existential) standpoint on the one hand and the intuitionistic (constructive) one on the other hand" (Bernays, 1946, p. 79).

In contrast to the medium of free becoming pertinent to A-time, the B-theoretic continuum is conceived as a rigid being (Weyl, 1925, p.18). It can be formalized by Cantor's concept of the arithmetic continuum of all real numbers, defined by the method of Dedekind cuts.[36]

9.2 Elementary B-Time Systems

The discussion of the affine group $\mathscr{B}$, defined by (13b) with the group relation

$$B(a,\tau)\,B(a',\tau') \;=\; B(aa',\tau+\tau'/a), \tag{34}$$

is formally analogous to the discussion of the affine group $\mathscr{A}$, but the conceptual implications are very different. The affine group $\mathscr{B}$ describes the homogeneous *tenseless B-time*. On the non-commutative algebra $\mathfrak{T}$ the affine group $\mathscr{B}$ is represented by the unitary operators

$$B(a,\tau) \;=\; V(\tau)\,X(a), \quad a>0, \quad \tau\in\mathscr{V}, \tag{35}$$

where $\{V(\tau)|\tau\in\mathscr{V}\}$ and $\{X(a)|a>0\}$ represent the time-translation and the scaling group, respectively. The fixed point algebra of the action of $\mathscr{B}$ on $\mathfrak{T}$ is given by

---

[35] Platonism is a form of realism which holds that mathematical entities exist "in themselves", independent of the human mind. In contrast, it is a *tenet* of intuitionism that no truth exists independent of our knowledge and that mathematical constructions are *intentional mental processes* carried out in time. For more details compare Tieszen (2005, part III).

[36] Cantor opposed the idea that the continuous is irreducible to the discrete and argued that "[time] can be conceived neither objectively as a substance, nor subjectively as a necessary *a priori* form of intuition but is nothing more than an auxiliary parameter for the description of movements in the material world" (Cantor, 1883, p. 573).

$$\mathfrak{B} \;:=\; \big\{F \in \mathfrak{T} \,\big|\, B(a,\tau)FB(a,\tau)^* = F \text{ for all } a > 0\,, \tau \in \mathscr{V}\big\}. \tag{36}$$

As in the case of A-time, the representation theory of the affine group implies that the W*-algebra $\mathfrak{B}$ generated by the affine group $\{B(a,\tau)|a > 0, \tau \in \mathscr{V}\}$ is the direct sum of three factors

$$\mathfrak{B} \;=\; \mathfrak{B}^- \oplus \mathfrak{B}^0 \oplus \mathfrak{B}^+, \tag{37}$$

where $\mathfrak{B}^0$ is the trivial W*-algebra of complex numbers $\mathbb{C}$. On the infinite-dimensional algebra $\mathfrak{B}^\pm$ the affine group $\mathscr{B}$ acts ergodically. The center of $\mathfrak{B}$ is generated by three orthogonal projectors Q^-, Q^0 and Q^+, defined by

$$Q^0\,\mathfrak{B}\,Q^0 \;=\; \mathfrak{B}^0, \quad Q^\pm\,\mathfrak{B}\,Q^\pm \;=\; \mathfrak{B}^\pm. \tag{38}$$

On the factors $\mathfrak{B}^+$ and $\mathfrak{B}^-$ the affine group $\mathscr{B}$ acts ergodically, so we can define the selfadjoint restrictions

$$\Lambda^\pm \;=\; \Lambda\big|_{\mathfrak{B}^\pm}, \quad S^\pm \;=\; S\big|_{\mathfrak{B}^\pm}. \tag{39}$$

with the positive and negative frequency operators,

$$\Lambda^+ \;>\; 0 \quad , \quad \Lambda^- \;<\; 0. \tag{40}$$

The projectors Q^+ and Q^- are classical observables allowing the following *Boolean classification*:

the projector Q_+ characterizes *spectral positivity,* (41a)
the projector Q_- characterizes *spectral negativity.* (41b)

A complex-valued signal with no negative-frequency components has in its time-representation an analytic continuation into a complex half-plane. Since the negative frequency components of a real-valued signal do not provide any information which is not already contained in the positive ones, such functions are frequently called *analytic signals* in communication theory and in optics (Gabor, 1946; Born and Wolf, 1959; Cohen, 1995).

9.3 Symmetry Breaking in the B-Time Description

In the B-time description time is static and homogeneous, and both time-translation and time-reversal symmetry are intact. On the elementary algebra $\mathfrak{B}^\pm$ the time evolution is given by the *group* $\{V^\pm(\tau)|\tau \in \mathscr{V}\}$ with

$$V^\pm(\tau) \;:=\; Q^\pm\,V(\tau)\,Q^\pm. \tag{42}$$

Since the B-time description is invariant under time translations it is impossible to define the concept of a Now in this description. Moreover in this description *the frequency-translation symmetry is broken* so that an elementary

B-time system does not allow to define a selfadjoint time operator. Nonetheless the symmetric (but not selfadjoint) restriction

$$T^{\pm} \;:=\; Q^{\pm}\, T\, Q^{\pm} \tag{43}$$

of the selfadjoint time operator T transforms covariantly under time translations, fulfilling the so-called *weak Weyl relation* (Schmüdgen, 1983),

$$e^{2\pi i \tau \Lambda^{\pm}}\, T^{\pm}\, e^{-2\pi i \tau \Lambda^{\pm}} \;=\; T^{\pm} + \tau\, \mathbf{1}. \tag{44}$$

9.4 Canonical Operators for the B-Time Description

The commutation relation (9b) implies the *affine commutation relation*

$$S\,\Lambda \,-\, \Lambda\, S \;=\; (i/2\pi)\,\Lambda \tag{45}$$

for the frequency operator Λ and the scale operator S, and therewith the affine commutation relation

$$S^{\pm}\,\Lambda^{\pm} \,-\, \Lambda^{\pm}\, S^{\pm} \;=\; (i/2\pi)\,\Lambda^{\pm} \tag{46}$$

for the operators restricted to the algebras $\mathfrak{B}^{+}$ and $\mathfrak{B}^{-}$. In the same way as in Sect. 8 we can define a selfadjoint logarithmic frequency operator $(\ln\Lambda)^{\pm}$ which fulfills with the scale operator $S^{\pm}$ the canonical commutation relations

$$S^{\pm}\,(\ln\Lambda)^{\pm} \,-\, (\ln\Lambda)^{\pm}\, S^{\pm} \;=\; (i/2\pi)\,\mathbf{1}. \tag{47}$$

10 Complementarity of A-Time and B-Time

Given two projection P and Q in an arbitrary W*-algebra, the projection

$$C(P,Q) \;:=\; (P \wedge Q) \vee (P \wedge Q^{\perp}) \vee (P^{\perp}\wedge Q) \vee (P^{\perp}\wedge Q^{\perp}), \tag{48}$$

measures the incompatibility of P and Q (Piron, 1966, p. 447). The following relations hold:

$$\mathbf{0} \le C(P,Q) \le \mathbf{1}, \tag{49a}$$

$$C(P,Q)\,P = P\,C(P,Q) \quad , \quad C(P,Q)\,Q = Q\,C(P,Q), \tag{49b}$$

$$C(P,Q) = \mathbf{1} \text{ if and only if } P\,Q = Q\,P, \tag{49c}$$

$$C(P,Q) = \mathbf{0} \text{ implies } P \wedge Q = P \wedge Q^{\perp} = P^{\perp}\wedge Q = P^{\perp}\wedge Q^{\perp} = \mathbf{0}. \tag{49d}$$

If $C(P,Q) = \mathbf{1}$, the projections P and Q are are called *compatible*, otherwise incompatible. If $C(P,Q) = \mathbf{0}$, there exists no normal state ρ such that both P and Q are truth-definite, that is, such that $\rho(P)$ and $\rho(Q)$ are both 0 or 1 (Raggio and Rieckers, 1983, prop. 2.2). Such projections are said to be *maximally incompatible.*

Let $P := P_+$ and $P^\perp = P_-$ be the projections (21) onto the invariant algebras $\mathfrak{A}_+$ and $\mathfrak{A}_-$ under the affine time group $\mathscr{A}$ of the A-time description, and let $Q := Q^+$ and $Q^\perp := Q^-$ be the projections (41) onto the invariant algebras $\mathfrak{B}^+$ and $\mathfrak{B}^-$ under the affine frequency group $\mathscr{B}$ of the B-time description. Then the relations $P_+ \wedge Q^+ = P_+ \wedge Q^- = P_- \wedge Q^+ = P_- \wedge Q^- = \mathbf{0}$ hold (Busch and Lahti, 1986, eq. 12). Therefore (P_+, Q^+), (P_+, Q^-), (P_-, Q^+) and (P_-, Q^-) are pairs of maximally incompatible projections implying that the descriptions of temporal phenomena in terms of the algebras $\mathfrak{A}_+$ and $\mathfrak{B}^+$, or $\mathfrak{A}_+$ and $\mathfrak{B}^-$, or $\mathfrak{A}_-$ and $\mathfrak{B}^+$, or $\mathfrak{A}_-$ and $\mathfrak{B}^-$ are maximally incompatible. That is, *A-time and B-time are complementary notions*, referring to *maximally incompatible and mutually exclusive complementary ways of description*, both providing *partial descriptions* of the phenomenon "time".

The complementarity of the unitarily equivalent pairs (P_+, Q^+), (P_+, Q^-), (P_-, Q^+) and (P_-, Q^-) are related to the unitary representation of the *infinite dihedral group* $\mathfrak{D}_\infty$ which is generated by the involutions $F = 2P_+ - 1$ and $G = 2Q^+ - 1$.[37] The irreducible unitary representations of $\mathfrak{D}_\infty$ are parametrized by an angle $\theta \in [0, \pi/2]$. In the irreducible representation with the parameter θ the projections P_+, Q^+ are given by

$$P_\theta := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad Q_\theta := \begin{pmatrix} \cos^2(\theta) & \cos(\theta)\sin(\theta) \\ \cos(\theta)\sin(\theta) & \sin^2(\theta) \end{pmatrix}, \tag{50}$$

The parameter θ is the value of a selfadjoint operator Θ, defined by

$$\sin(\Theta) := |\, P_+ - Q^+|. \tag{51}$$

It describes the operator-valued angle between the projections P_+ and Q^+. The operator Θ generates the center of the algebra spanned by the the projectors P_+ and Q^+ (see Takesaki, 1979, p. 306–308). The scale operator S commutes with $P_\pm$ and with $Q^\pm$ and is related to the angle operator Θ by

$$\tan(\Theta) := e^{-2\pi^2 S}. \tag{52}$$

The complementarity of the projectors $P_\pm$ and $Q^\pm$ implies that descriptions in terms of tensed time and descriptions in terms of tenseless time cannot be captured in a single Boolean description. Elementary A-time systems with the time operators T_+ and T_- realize a superselection rule between past and future which is traditionally presupposed for the description of input–output systems. They can be described in terms of the Hardy space $\widehat{\boldsymbol{H}}^2(\mathbb{C}^\pm)$

---

[37] A group generated by two elements F and G of period 2 is called *dihedral.* Every dihedral group is generated by a rotation R and a reflection. If the rotation is a rational multiple of a full rotation, then there is some integer n such that R^n is the identity, and we have a *finite dihedral group* $\mathfrak{D}_n$of order $2n$. If the rotation is not a rational multiple of a full rotation, then there is no such n and the resulting group has infinitely many elements and is called the *infinite dihedral group* $\mathfrak{D}_\infty$.

whose elements are the inverse Fourier transforms of square-integrable functions $t \mapsto f(t)$ which vanish on the half-axis $\mathbb{R}^{\mp}$, $P_{\mp}f = 0$. Similarly, elementary B-time systems with the frequency operators Λ^{+} and Λ^{-} realize a superselection rule between spectrally positive and spectrally negative processes. They can be described in terms of the Hardy space $\boldsymbol{H}^2(\mathbb{C}^{\pm})$ whose elements are the Fourier transforms of square-integrable functions $\lambda \mapsto \widehat{f}(\lambda)$ which vanish on the half-axis $\mathbb{R}^{\mp}$, $Q^{\mp}\widehat{f} = 0$. The maximal incompatibility of the projections $P_{\pm}$ and $Q^{\pm}$ implies that the descriptions in terms of the Hardy spaces $\widehat{\boldsymbol{H}}^2(\mathbb{C}^{\pm})$ and $\boldsymbol{H}^2(\mathbb{C}^{\pm})$ are complementary in the sense of maximal incompatibility.

The *A-time* characterized by the affine time group $\mathscr{A}$ corresponds to the tensed time that we encounter in experience, but it does not reflect a feature of some fundamental law of physics. The *Now* is the characteristic feature of A-time, requiring that the time-translation and the time-reversal symmetry of fundamental physical principles are broken. Tensed structures with a preferred point of reference arise not only in subjective experience but also in experimental science. The tensed A-time distinguishes past and future in such a way that holistic correlations between past and future are impossible. Every evaluation of an experiment depends on this sharp distinction between a *past stimulus* and a *future response.* Without these distinctions it would not be possible to test, verify or falsify theories.

The *B-time* characterized by the affine frequency group $\mathscr{B}$ corresponds to the homogeneous tenseless time that we encounter in physics where all basic principles remain the same at all times. Furthermore, the homogeneity of B-time is necessary and sufficient for the validity of the *conservation law of energy.* In a description in terms of B-time we may have holistic correlations between the domains which in the A-time description are called “past” and “future”.

Yet, physics cannot get along with B-time alone. Since all time instants of physical B-time are equivalent, the first principles of physics on their own cannot explain experiments. The specification of the initial conditions for an experiment requires the introduction of A-time into physics.

11 Extended Time Systems

11.1 Possible Descriptions of Extended Time Systems

Although one can consider the atemporal material and mental domains separately (see Sect. 5.2), our ultimate goal is to describe how time is affiliated with both domains. This can be achieved by extending the relevant symmetries of the temporal domain to the whole universe of discourse. Depending on the chosen description of the temporal domain, we can distinguish the following four descriptions:

$$\text{fully holistic description:} \quad \mathfrak{L} = \mathfrak{M} \mathbin{\bar{\otimes}} \mathfrak{T} \mathbin{\bar{\otimes}} \mathfrak{N}, \tag{53a}$$

$$\text{extended A-time description:} \quad \mathfrak{L}_{\mathrm{A}} = \mathfrak{M} \mathbin{\bar{\otimes}} \left\{ \mathfrak{A}_{-} \oplus \mathfrak{A}_{0} \oplus \mathfrak{A}_{+} \right\} \mathbin{\bar{\otimes}} \mathfrak{N}, \tag{53b}$$

$$\text{extended B-time description:} \quad \mathfrak{L}_{\mathrm{B}} = \mathfrak{M} \mathbin{\bar{\otimes}} \left\{ \mathfrak{B}^{-} \oplus \mathfrak{B}^{0} \oplus \mathfrak{B}^{+} \right\} \mathbin{\bar{\otimes}} \mathfrak{N}, \tag{53c}$$

$$\text{classical-time description:} \quad \mathfrak{L}_{\mathrm{C}} = \mathfrak{M} \mathbin{\bar{\otimes}} \mathfrak{C} \mathbin{\bar{\otimes}} \mathfrak{N}. \tag{53d}$$

In the fully holistic description the center $\mathfrak{Z}(\mathfrak{L})$ is trivial (i.e. it consists only of multiples of the identity operator). In a generic state the material, the temporal and the mental domain are mutally holistically correlated so that material, temporal and mental phenomena cannot be clearly distinguished.

In the description with a classical time operator, the center $\mathfrak{Z}_{\mathrm{C}}(\mathfrak{L}_{\mathrm{C}})$ equals essentially the commutative algebra $\mathfrak{C}$ generated by the classical parametric time operator excluding holistic time phenomena. In particular, in this description the material and the mental domain are not holistically correlated via the temporal domain.

A-time and B-time descriptions are intermediate between the fully holistic and the classical time description. The center $\mathfrak{Z}_{\mathrm{A}}(\mathfrak{L}_{\mathrm{A}})$ of the A-time description is generated by the classical observables (18), where the projection P_{-} selects the past and the projection P_{+} selects the future. The center $\mathfrak{Z}_{\mathrm{B}}(\mathfrak{L}_{\mathrm{B}})$ of the B-time description is generated by the classical observables (38), where the projection Q^{-} selects spectrally negative processes and the projection Q^{+} selects spectrally positive processes. The smallest W*-subalgebra $\mathfrak{L}_{\mathrm{A}} \vee \mathfrak{L}_{\mathrm{B}}$ which contains both the algebra $\mathfrak{L}_{\mathrm{A}}$ of the A-time description and the algebra $\mathfrak{L}_{\mathrm{B}}$ of the B-time description is not a factor but has a nontrivial center generated by the operator $\mathbf{1} \otimes \Theta \otimes \mathbf{1}$, where Θ is the operator-valued angle (51) between the projections P_{+} and Q^{+}.

11.2 Time–Matter Systems

Time plays an important role in the description of the material domain. However, the fundamental laws of physics are invariant under time-translations and under time-reversal. This implies that notions like past and future, or irreversibility, do not appear in fundamental physics. What is called "time" in fundamental physics is just a *correlation parameter*.[38] Moreover, only in physics, the temporal correlation parameter is linked with the spatial correlation parameters (e.g. via the Lorentz or the Galilei group).

It is an amazing fact that (for a fixed observer) all material phenomena can be correlated *linearly*. We formalize this result in terms of a one-parameter automorphism group $\{\alpha_\tau | \tau \in \mathbb{R}\}$ of the algebra $\mathfrak{M}$ of the atemporal material domain. Since every automorphism of $\mathfrak{M}$ is inner, it can be implemented by a unitary group $\{e^{i\tau H/\hbar} | \, \tau \in \mathbb{R}\}$, where the selfadjoint operator H is referred to

---

[38] In classical physics it is well-known that by a canonical transformation it is possible to generate an atemporal representation (see Synge, 1960, pp. 100 ff). In such a framework there is no reference to time. A similar reformulation is also possible in quantum mechanics.

as the Hamiltonian of the material domain. For each $\tau \in \mathbb{R}$ this automorphism of $\mathfrak{M}$ is given by

$$\alpha_\tau(M) \;=\; e^{2\pi i \tau H/h} M \, e^{-2\pi i \tau H/h} \in \mathfrak{M} \quad \text{for every} \;\; M \in \mathfrak{M}. \tag{54}$$

The *auxiliary parameter* τ describes *correlations between material events*, and should not be confused with time as experienced. According to Gödel "the real idea behind time is causation; the time structure of the world is just its causal structure" (conversation on 25. 11. 1975, reported by Wang, 1995, p. 229).

In extended time–matter system, *time is the expectation value of the extended time operator* – in the A-time description the expectation value of $\boldsymbol{T}_\pm = \mathbf{1} \otimes T_\pm \otimes \mathbf{1}$ and in the B-time description the expectation value of $\boldsymbol{T}^\pm = \mathbf{1} \otimes T^\pm \otimes \mathbf{1}$. In the A-time and B-time description the auxiliary real-valued parameter τ is correlated to A-time via the parametric unfoldings (27) and (44), respectively:

$$\text{A-time:} \quad \tau \;\rightarrow\; \boldsymbol{T}_\pm(\tau) \;=\; \boldsymbol{T}_\pm \,+\, \tau \mathbf{1} \;\in\; \mathfrak{M} \,\bar{\otimes}\, \mathfrak{A}_\pm \,\bar{\otimes}\, \mathfrak{N}, \tag{55a}$$

$$\text{B-time:} \quad \tau \;\rightarrow\; \boldsymbol{T}^\pm(\tau) = \boldsymbol{T}^\pm \,+\, \tau \mathbf{1} \;\in\; \mathfrak{M} \,\bar{\otimes}\, \mathfrak{B}^\pm \,\bar{\otimes}\, \mathfrak{N}. \tag{55b}$$

If the state of the extended system is given by the state functional ρ, then the parameter τ is given by

$$\text{A-time:} \quad \tau \;=\; \rho\{\boldsymbol{T}_\pm(\tau)\} - \rho\{\boldsymbol{T}_\pm\}\,, \tag{56a}$$

$$\text{B-time:} \quad \tau \;=\; \rho\{\boldsymbol{T}^\pm(\tau)\} - \rho\{\boldsymbol{T}^\pm\}. \tag{56b}$$

It is convenient to write $\tau = \bar{t} - \bar{t}_0$, where $\bar{t}$ is the mean value of the relevant time operator at the parameter value τ, and $\bar{t}_0$ is the reference point for the time calibration,

$$\text{A-time:} \quad \bar{t} \;:=\; \rho\{\boldsymbol{T}_\pm(\tau)\} \;\; \text{and} \;\; \bar{t}_0 \;:=\; \rho\{\boldsymbol{T}_\pm\}\,, \tag{57a}$$

$$\text{B-time:} \quad \bar{t} \;:=\; \rho\{\boldsymbol{T}^\pm(\tau)\} \;\; \text{and} \;\; \bar{t}_0 \;:=\; \rho\{\boldsymbol{T}^\pm\}. \tag{57b}$$

If the temporal domain is neither correlated with the mental nor with the material domain (that is, if ρ is a product state), the quantity $\bar{t}_0$ only depends on the state of the temporal domain. In this case we can, without loss of generality, put $\bar{t}_0 = 0$, so that $\tau = \bar{t}$. This is no longer possible for entangled systems.

In terms of the physically relevant mean time $\bar{t}$ the expectation value of a material observable M is given by

$$m(\bar{t}) \;=\; \rho\{M(\bar{t} - \bar{t}_0) \otimes \mathbf{1} \otimes \mathbf{1}\}. \tag{58}$$

Here $\bar{t}$ can be interpreted as the mean of a random time with the variance

$$\text{A-time:} \quad \sigma^2 \;=\; \rho\big\{(\mathbf{1} \otimes \boldsymbol{T}_\pm \otimes \mathbf{1})^2\big\} - \rho\big\{\mathbf{1} \otimes \boldsymbol{T}_\pm \otimes \mathbf{1}\big\}^2\,, \tag{59a}$$

$$\text{B-time:} \quad \sigma^2 \;=\; \rho\big\{(\mathbf{1} \otimes \boldsymbol{T}^\pm \otimes \mathbf{1})^2\big\} - \rho\big\{\mathbf{1} \otimes \boldsymbol{T}^\pm \otimes \mathbf{1}\big\}^2. \tag{59b}$$

For holistically correlated systems both the reference time $\bar{t}_0$ and the variance σ^2 depend on the state of the material and/or the mental domain.

11.3 Extended A-Time Descriptions

11.3.1 Representation of the “Medium of Free Becoming”

The ordering in the temporal A-domain is governed by the group structure of the continuum $\mathbb{R}$, represented by the one-parameter unitary group $\mathscr{V}$. Weyl’s “medium of becoming” can be represented by a structure which was introduced into scattering theory by Lax and Phillips (1967). An orthogonal *Lax–Phillips structure* deals with a unitary one-parameter group $\{\boldsymbol{V}(\tau)|\tau \in \mathbb{R}\}$, acting on some Hilbert space $\boldsymbol{\mathcal{H}}$ with two closed subspaces $\boldsymbol{\mathcal{N}}_-$ and $\boldsymbol{\mathcal{N}}_+$, distinguished by the following double K-structure:

$$\boldsymbol{V}(\tau)\,\boldsymbol{\mathcal{N}}_+ \;\subset\; \boldsymbol{\mathcal{N}}_+ \quad \text{for } \tau > 0, \quad \boldsymbol{V}(\tau)\,\boldsymbol{\mathcal{N}}_- \;\subset\; \boldsymbol{\mathcal{N}}_+ \quad \text{for } \tau < 0, \tag{60a}$$

$$\wedge_{\tau\in\mathbb{R}} \boldsymbol{V}(\boldsymbol{\tau})\,\boldsymbol{\mathcal{N}}_+ \;=\; \{0\} \;=\; \wedge_{\tau\in\mathbb{R}} \boldsymbol{V}(\boldsymbol{\tau})\,\boldsymbol{\mathcal{N}}_-, \tag{60b}$$

$$\vee_{\tau\in\mathbb{R}} \boldsymbol{V}(\tau)\,\boldsymbol{\mathcal{N}}_+ \;=\; \boldsymbol{\mathcal{N}} \;=\; \vee_{\tau\in\mathbb{R}} \boldsymbol{V}(\tau)\,\boldsymbol{\mathcal{N}}_-, \tag{60c}$$

$$\boldsymbol{\mathcal{N}}_+ \perp \boldsymbol{\mathcal{N}}_- . \tag{60d}$$

In the Lax–Phillips theory the subspaces $\boldsymbol{\mathcal{N}}_-$, $\boldsymbol{\mathcal{N}}_+$, and $\boldsymbol{\mathcal{N}}_0$ are referred to as the *incoming subspace*, the *outgoing subspace*, and the *interaction subspace*, respectively. If $\boldsymbol{\mathcal{N}}_0 = \{\boldsymbol{0}\}$, then the Lax–Phillips system is called *trivial.*

Every Lax–Phillips structure has a *translation representation* in terms of $\mathcal{K}$-valued functions $t \mapsto \Xi(t)$, where $\mathcal{K}$ is an auxiliary Hilbert space,

$$\boldsymbol{\mathcal{H}} \;=\; \boldsymbol{L}^2(\mathbb{R}, dt) \otimes \mathcal{K} \;=\; \boldsymbol{\mathcal{N}}_- \oplus \boldsymbol{\mathcal{N}}_0 \oplus \boldsymbol{\mathcal{N}}_+, \tag{61a}$$

$$\boldsymbol{\mathcal{N}}_\pm \;=\; \boldsymbol{L}^2(\mathbb{R}^\pm, dt) \otimes \mathcal{K}, \quad \boldsymbol{\mathcal{N}}_0 \;=\; \mathbb{C} \otimes \mathcal{K}. \tag{61b}$$

On the Hilbert space $\boldsymbol{L}^2(\mathbb{R}, dt) \otimes \mathcal{K}$ the unitary group $\{\boldsymbol{V}(\tau)|\tau \in \mathbb{R}\}$ acts as a *time-translation group*,

$$\{\boldsymbol{V}(\tau)\,\Xi\}(t) \;=\; \{V(\tau) \otimes \boldsymbol{1})\,\Xi\}(t) \;=\; \Xi(t-\tau), \quad \Xi \in \boldsymbol{L}^2(\mathbb{R}, dt) \otimes \mathcal{K}, \tag{62}$$

while the three elementary representations of the affine group $\mathscr{A} = \{A(a,b)|a > 0, b \in \mathscr{U}\}$ are represented by

$$\big\{\boldsymbol{A}_0(a,b)\,\boldsymbol{1}_\mathbb{C} \otimes \Psi\big\}(z) = \boldsymbol{1}_\mathbb{C}(z^a) \otimes \Psi, \quad z \in \mathbb{C}, \tag{63a}$$

$$\big\{\boldsymbol{A}_\pm(a,b)\,\Phi_\pm \otimes \Psi\big\}(t) = e^{2\pi i b t}\,\sqrt{a}\;\Phi_\pm(at) \otimes \Psi \;\;,\;\; \Phi_\pm \in \boldsymbol{L}^2(\mathbb{R}^\pm, dt), \tag{63b}$$

where $\boldsymbol{1}_\mathbb{C}$ is the identity function in $\mathbb{C}$, $\boldsymbol{1}_\mathbb{C}(z) = z$, and $\Psi \in \mathcal{K}$. The W*-algebras $\mathfrak{B}(\boldsymbol{\mathcal{N}}_\pm)$ and $\mathfrak{B}(\boldsymbol{\mathcal{N}}_\mathbf{0})$ are related to the W*-algebras $\mathfrak{A}_\pm$ and $\mathfrak{A}_0$ by

$$\mathfrak{B}\{\boldsymbol{\mathcal{N}}_\pm\} \;=\; \mathfrak{A}_\pm \,\bar{\otimes}\, \mathfrak{K}, \quad \mathfrak{B}\{\boldsymbol{\mathcal{N}}_\mathbf{0}\} \;=\; \mathfrak{A}_0 \,\bar{\otimes}\, \mathfrak{K}, \tag{64}$$

where $\mathfrak{A}_0 = \mathfrak{B}(\mathbb{C})$ and $\mathfrak{K} = \mathfrak{B}(\mathcal{K}) \subseteq \mathfrak{M} \,\bar{\otimes}\, \mathfrak{N}$.

11.3.2 Prediction and Retrodiction of Signals

If the Hilbert space $\boldsymbol{\mathcal{H}}$ is generated by a family of square-integrable information-carrying $\mathcal{K}$-valued signals $\{\boldsymbol{s}_\tau | \tau \in \mathbb{R}\}$ with $\boldsymbol{s}_\tau = \boldsymbol{V}(\tau)\boldsymbol{s}_0$ and $\boldsymbol{s}_0 \in \boldsymbol{\mathcal{H}}$, we can define the closed subspace

$$\boldsymbol{\mathcal{H}}_\tau \;:=\; \overline{\mathrm{span}}\,\{\boldsymbol{s}(\tau') | \tau' \le \tau\}. \tag{65}$$

where $\overline{\mathrm{span}}$ stands for the closed linear hull. The subspace $\boldsymbol{\mathcal{H}}_\tau$ can be conceived as representing the *information* available at A-time τ. As τ increases, the Hilbert spaces $\boldsymbol{\mathcal{H}}_\tau$ form a never decreasing family, so that the limiting spaces $\boldsymbol{\mathcal{H}}_{-\infty}$ and $\boldsymbol{\mathcal{H}}_{+\infty}$ exist. To exclude trivial situations we assume that the process $\tau \mapsto \boldsymbol{s}(\tau)$ is purely nondeterministic in the sense that the remote past $\boldsymbol{\mathcal{H}}_{-\infty} = \wedge_\tau \boldsymbol{\mathcal{H}}_\tau$ contains only the zero element of $\boldsymbol{\mathcal{H}}$.

We now define a subspace $\boldsymbol{\mathcal{H}}_-$ spanned by the past signals $\{\boldsymbol{s}(\tau) | \tau \le 0\}$ and a subspace $\boldsymbol{\mathcal{H}}_+$ spanned by the future signals $\{s(\tau) | \tau \ge 0\}$,

$$\boldsymbol{\mathcal{H}}_- \;:=\; \vee_{\tau \le 0}\, \boldsymbol{\mathcal{H}}_\tau, \quad \boldsymbol{\mathcal{H}}_+ \;:=\; \vee_{\tau \ge 0}\, \boldsymbol{\mathcal{H}}_\tau. \tag{66}$$

Therewith we get

$$\boldsymbol{V}(\tau)\,\boldsymbol{\mathcal{H}}_- \subseteq \boldsymbol{\mathcal{H}}_- \quad \text{for } \tau < 0, \quad \boldsymbol{V}(\tau)\,\boldsymbol{\mathcal{H}}_+ \subseteq \boldsymbol{\mathcal{H}}_+ \quad \text{for } \tau > 0. \tag{67}$$

The full history is given by $\boldsymbol{\mathcal{H}}_- \vee \boldsymbol{\mathcal{H}}_+$, but because the subspaces $\boldsymbol{\mathcal{H}}_-$ and $\boldsymbol{\mathcal{H}}_+$ overlap, they are not orthogonal. Let $\boldsymbol{P}_\pm$ be the orthogonal projection of $\boldsymbol{\mathcal{H}}$ onto $\boldsymbol{\mathcal{H}}_\pm$, then the Hilbert spaces

$$\boldsymbol{\mathcal{H}}_{+/-} \;=\; \boldsymbol{P}_-\,\boldsymbol{\mathcal{H}}_+, \quad \boldsymbol{\mathcal{H}}_{-/+} \;=\; \boldsymbol{P}_+\,\boldsymbol{\mathcal{H}}_-. \tag{68}$$

can be interpreted as conditional expectations (in the linear sense), given that the process $\{\boldsymbol{s}_\tau | \pm \tau \le 0\}$ is known. That is, the subspace $\boldsymbol{\mathcal{H}}_{+/-}$ is the smallest subspace containing all the information from the past $\boldsymbol{\mathcal{H}}_-$ needed for predicting the future of the process in terms of a linear expression of the past of the process. Accordingly $\boldsymbol{\mathcal{H}}_{+/-}$ is called the *forward predictor space* (in the linear sense). Similarly $\boldsymbol{\mathcal{H}}_{-/+}$ is called the *backward predictor space* (see Lindquist and Picci, 1991, and references given there).

11.3.3 Classical Correlations Between Past and Future

The so-called frame space

$$\boldsymbol{\mathcal{N}}_0 \;:=\; \boldsymbol{\mathcal{H}}_{+/-} \vee \boldsymbol{\mathcal{H}}_{-/+}, \tag{69}$$

corresponds to the interaction subspace of the Lax–Phillips theory. *It acts as the state space for $t = 0$ and contains all classical correlations between past and future.*

The part of the space $\boldsymbol{\mathcal{H}}_-$ of past signals which is orthogonal to the space $\boldsymbol{\mathcal{H}}_+$ is given by $\boldsymbol{\mathcal{N}}_-$, and the part of the space $\boldsymbol{\mathcal{H}}_+$ of future signals which is orthogonal to the past $\boldsymbol{\mathcal{H}}_-$ is given by $\boldsymbol{\mathcal{N}}_+$, where

$$\boldsymbol{\mathcal{N}}_{\mp} \;:=\; \boldsymbol{\mathcal{H}}_{\mp} \wedge (\boldsymbol{\mathcal{H}}_{\pm})^{\perp} = \boldsymbol{\mathcal{H}}_{\mp} \ominus \boldsymbol{\mathcal{H}}_{\pm/\mp}, \quad \boldsymbol{\mathcal{N}}_+ \perp \boldsymbol{\mathcal{N}}_- . \tag{70}$$

This relation leads to the orthogonal Lax–Phillips decomposition (61a).

The one-parameter group $\{\boldsymbol{V}(\tau)|\tau\in\mathbb{R}\}$ maps the incoming subspace $\boldsymbol{\mathcal{N}}_-$ into all three subspaces, so that the past influences the present and the future. Furthermore the group $\{\boldsymbol{V}(\tau)|\tau\in\mathbb{R}\}$ maps the interaction subspace $\boldsymbol{\mathcal{N}}_0$ into itself and into the outgoing subspace $\boldsymbol{\mathcal{N}}_+$, so that the present influences only the present and the future. Since the group $\{\boldsymbol{V}(\tau)|\tau \in \mathbb{R}\}$ maps outgoing subspace $\boldsymbol{\mathcal{N}}_+$ into itself, the future has no influence on the present or the past. In this sense, the one-parameter group $\{\boldsymbol{V}(\tau)|\tau\in\mathbb{R}\}$ decribes the *flow of A-time.*

The incoming subspace $\boldsymbol{\mathcal{N}}_-$ is invariant under the action of $\boldsymbol{V}(\tau)$ for $\tau \geq 0$, and the outgoing subspace $\boldsymbol{\mathcal{N}}_+$ is invariant under the action of $\boldsymbol{V}^*(\tau)$ for $\tau \geq 0$. Thus, the A-time dynamics respects the superselection rule between past and future which prevents holistic correlations between past and future. The three elementary representations are characterized by the three time operators $\boldsymbol{T_-} := \mathbf{1} \otimes T_- \otimes \mathbf{1}$, $\boldsymbol{T_0} := \mathbf{1} \otimes T_0 \otimes \mathbf{1}$, and $\boldsymbol{T_+} := \mathbf{1} \otimes T_+ \otimes \mathbf{1}$. In the extended description the A-time-concepts "past", "now" and "future" of the temporal domain are now represented by the *spaces* $\boldsymbol{\mathcal{N}}_-$, $\boldsymbol{\mathcal{N}}_0$ and $\boldsymbol{\mathcal{N}}_+$, where according to the Lax-Phillips decomposition (61a):

the space $\boldsymbol{\mathcal{N}}_-$ represents the *past of the extended system*,

$$\text{with } \langle \Phi_- | \boldsymbol{T_-} \Phi_- \rangle < 0 \;\text{ for }\; \Phi_- \in \boldsymbol{\mathcal{N}}_-, \tag{71a}$$

the space $\boldsymbol{\mathcal{N}}_0$ represents the *Now of the extended system*,

$$\text{with } \langle \Phi_0 | \boldsymbol{T_0}\, \Phi_0 \rangle = 0 \;\text{ for }\; \Phi_0 \in \boldsymbol{\mathcal{N}}_0, \tag{71b}$$

the space $\boldsymbol{\mathcal{N}}_+$ represents the *future of the extended system*,

$$\text{with } \langle \Phi_+ | \boldsymbol{T_+} \Phi_+ \rangle > 0 \;\text{ for }\; \Phi_+ \in \boldsymbol{\mathcal{N}}_+. \tag{71c}$$

11.3.4 The Extended Now

The value of the time operator $\boldsymbol{T_0}$ in the space $\boldsymbol{\mathcal{N}}_0$ of the extended Now is the punctual now (21b). If the extended Now is generated by an information-carrying signal $\tau \mapsto \boldsymbol{s}(\tau)$, then the extended Now $\boldsymbol{\mathcal{N}}_0$ consists not only of the initial value $\boldsymbol{s}(0)$, but is spanned by $\boldsymbol{s}(0)$ together with all its one-sided derivatives at point $\tau = 0$.[39] Therefore it contains the minimal amount of

---

[39] If a signal $\tau \mapsto \boldsymbol{s}(\tau)$ is sufficiently smooth, its forward and backward derivatives exist. The backward differential operator D_- is defined by $\{D_-\boldsymbol{s}\}(\tau) = \lim_{\varepsilon\to 0^+} \{\boldsymbol{s}(\tau) - \boldsymbol{s}(\tau-\varepsilon)\}/\varepsilon$, while the forward differential operator D_+ is defined by $\{D_+\boldsymbol{s}\}(\tau) = \lim_{\varepsilon\to 0^+} \{\boldsymbol{s}(\tau+\varepsilon) - \boldsymbol{s}(\tau)\}/\varepsilon$. The backward derivatives

information for a linear prediction and a linear retrodiction of the process $\tau \mapsto \boldsymbol{s}(\tau)$, $\tau \in \mathbb{R}$.

The fact that the expectation value of the time operator $\boldsymbol{T_0}$ is always zero does not imply that the duration of the extended Now vanishes. Lax and Phillips (1978) introduced the selfadjoint operator

$$\boldsymbol{\Delta} := \operatorname*{s\text{-}lim}_{\varepsilon \to 0} \int_{-\infty}^{\infty} e^{-\varepsilon|\tau|} \boldsymbol{V}(-\tau)\, \boldsymbol{P}_0\, \boldsymbol{V}(\tau)\, d\tau \ \text{ with } \ \boldsymbol{\Delta V}(\tau) = \boldsymbol{V}(\tau)\boldsymbol{\Delta}, \quad (72)$$

as *time delay operator* of the subspace $\boldsymbol{\mathcal{N}}_0$, where $\boldsymbol{P}_0$ is the projection of $\boldsymbol{\mathcal{H}}$ onto $\boldsymbol{\mathcal{N}}_0$. If $\rho_0 := \rho \restriction \mathfrak{B}\{\boldsymbol{\mathcal{N}_0}\}$ is the restriction of the state functional ρ of the temporal system onto the algebra $\mathfrak{B}\{\boldsymbol{\mathcal{N}_0}\}$, then $\rho_0(\boldsymbol{P}_0)$ can be interpreted as the probability of finding the system in the subspace $\boldsymbol{\mathcal{N}}_0$.

In the context of Husserl's phenomenology the extended Now represents the "inner time of consciousness", constituted by "retention", "primal impression", and "protention" (Husserl, 1966). Retention and protention are not temporal intervals.[40] According to Husserl (1983, §81) perception has an intrinsic extendedness which cannot be measured by any physical means. In our representations this is expressed by the fact that intrinsic duration is not an interval of the spectrum of a time operator, but a property of the space $\boldsymbol{\mathcal{N}}_0$.

11.3.5 A-Time in Physics: Stimulus/Response Experiments

If we can discriminate between a material object system and its environment, then the atemporal material domain can be represented by the tensor product $\mathfrak{M} = \mathfrak{M}_{\text{obj}} \bar{\otimes} \mathfrak{M}_{\text{env}}$, where the W*-algebra $\mathfrak{M}_{\text{obj}}$ refers to the object system and the W*-algebra $\mathfrak{M}_{\text{env}}$ describes the enviroment. Since *every* subsystem of the material world is interacting with the rest of the world, the restriction of the dynamics $\tau \to \alpha_\tau$ to $\mathfrak{M}$ cannot any longer be an automorphism. If the interaction of the object system with its environment is weak,[41] the result-

---

$\boldsymbol{s}_-^{(n)}(0) := \{D_-^n \boldsymbol{s}\}(0)$ $(n = 0, 1, 2, \cdots)$ at the puntual now $t = 0$ incorporate information necessary for a partial *retrodiction* of the signal $\tau \mapsto \boldsymbol{s}(\tau)$. Similarly, the forward derivatives $\boldsymbol{s}_+^{(n)}(0) := \{D_-^n \boldsymbol{s}\}(0)$ $(n = 0, 1, 2, \cdots)$ at $t = 0$ allow a partial *prediction* of the signal.

[40] Retention is not to be confused with remembrance (which refers to the faculty of storing what has once been present to consciousness). In our representation remembrance is to be described in terms of the past-space $\boldsymbol{\mathcal{N}}_-$. Similarly, protention is not to be confused with expectation which is to be described in terms of the future-space $\boldsymbol{\mathcal{N}}_+$.

[41] More precisely: Assuming that the object system and its environment are initially uncorrelated and only weakly interacting, one gets first the completely positive semigroup $\beta_\tau(M_{\text{obj}}) := \alpha_\tau(M_{\text{obj}} \otimes \mathbf{1})$, $M_{\text{obj}} \in \mathfrak{M}_{\text{obj}}$, $\tau \geq 0$, a so-called *dynamical semigroup* (see Davies, 1976, chapt. 9). Since every dynamical semigroup can be dilated to a semigroup $\{\gamma_\tau | \tau \geq 0\}$ of normal unital *-endomorphisms (see Bhat, 1996), we can assume without loss of generality that β_τ is an endomorphism of the algebra $\mathfrak{M}_{\text{obj}}$.

ing retarded irreversible Markov dynamics can be described by a semigroup $\{\beta_\tau | \tau \geq 0\}$ of endomorphisms of the algebra $\mathfrak{M}_{\mathrm{obj}}$.

The dynamics $\tau \mapsto \beta_\tau$ is constitutive for a proper description of stimulus/response experiments. The intentional character of experiments (see Sect. 3.3) presumes a distinction between past and future, implying a description in terms of A-time given by a one-parameter semigroup of endomorphisms of the algebra $\mathfrak{M} \mathbin{\bar{\otimes}} (\mathfrak{A}_- \oplus \mathfrak{A}_+)$. In this case the parameter τ of the material dynamics can be correlated with the expectation value (57a) of the *A-time operator* $T_\pm$. In this description the flow of time is given by the semigroup $\{\gamma_\tau | \tau \geq 0\}$ of the endomorphisms defined by

$$\gamma_\tau(M \otimes Y) := (\beta \otimes \theta)_\tau (M \otimes Y) = \beta_\tau(M) \otimes \theta_\tau(Y) \ , \quad M \in \mathfrak{M} \, , \ Y \in \mathfrak{A}_+ \ , \qquad (73)$$

where β_t and θ_t refer to the material and the temporal domain, respectively.

The irreversibility of the map $\tau \mapsto \gamma_\tau$ is indispensable for the description of the appearance of facts. Without recourse to thermodynamic considerations, the irreversibility of the map $\tau \mapsto \beta_\tau$ is determined by the passivity of the transfer function of the input/output system in question (König and Tobergte, 1963; Dolph, 1963), which in turn is determined by the spectral properties of the Lax–Phillips semigroup $\{\boldsymbol{P}_0 \boldsymbol{V}(\tau) \boldsymbol{P}_0 | \tau \geq 0\}$.

The Fourier–Laplace transform $\int_0^\infty e^{-2\pi i z \tau} \gamma_\tau \, d\tau$ exists and is holomorphic in the open half-plane $\mathrm{Im}(z) < 0$. This allows us to use the powerful tools of the theory of Hardy functions to discuss *causal functions.*[42] Here causality means that a response should never take place before the stimulus that creates it. This time-honored principle is not necessitated by any fundamental physical law, but substantiated by the success of traditional phenomenological physical and engineering descriptions. With respect to mental systems, conscious perception and cognition also presuppose the usual forward direction of time, implying a memory of the past, but no anticipation of the future.

11.4 B-Time in Physics: Matter in Equilibrium

In the atemporal description of matter there is no genuine time so that the concept of a "stationary state" is not defined. Rovelli (1993) proposed to define a physical time by selecting an appropriate state and to declare this state as stationary. An atemporal definition of an equilibrium state is possible by requiring the stability with respect to small local perturbations which leads to the so-called KMS-condition, characteristic for thermal equilibrium states (Haag *et al.*, 1974).

---

[42] A function $\tau \mapsto f(\tau)$ is called causal if it is zero for $\tau < 0$. A function f defined on $\mathbb{C}^-$ is said to be a Hardy function if it is holomorphic on $\mathbb{C}^-$ and $\sup_{a>0} \int_{\mathbb{R}} |f(x + iy)|^2 dy < \infty$. The theory of Hardy spaces provides the proper setting for the discussion of stimulus/response systems and for the prediction theory of stochastic processes.

In elementary quantum mechanics a thermal equilibrium state ρ_β is defined by $\rho_\beta(M) = \mathrm{tr}(e^{-\beta H}M)/\mathrm{tr}(e^{-\beta H})$, $M \in \mathfrak{M}$, $\beta > 0$. This definition implies that the Hamiltonian H is bounded from below (so that we may assume that $H \geq 0$) and has a purely discrete spectrum. In this case the one-parameter group $\alpha_\tau(X) = e^{i\tau H/\hbar}Xe^{-i\tau H/\hbar}$ can be analytically continued to imaginary times, so that the function $z \mapsto \rho_\beta\{X\,\alpha_z(Y)\}$ is analytic in the open strip $\{z\,|\,0 < \mathrm{Im}(z) < \beta\}$, continuous and bounded on the closed strip $\{z\,|\,0 \leq \mathrm{Im}(z) \leq \beta\}$. In addition the so-called Kubo-Martin-Schwinger (KMS) boundary condition

$$\rho_\beta\{\alpha_\tau(X)\,Y\} \;=\; \rho_\beta\{Y\alpha_{\tau+i\beta}(X)\}, \quad \tau \in \mathbb{R}, \quad \beta \geq 0, \tag{74}$$

is satisfied for all $X, Y \in \mathfrak{M}$, implying that the state ρ_β is stationary in the sense of $\rho_\beta\{\alpha_\tau(X) = \rho_\beta(X)$. The KMS-condition characterizes thermal equilibrium states not only in elementary quantum mechanics but also in many cases where the Hamiltonian has a continuous spectrum (see Haag *et al.*, 1967).

On the basis of a KMS-equilibrium state one can define the one-parameter modular group of automorphisms (called the *modular group*). This "modular dynamics" can be used to define a *state-dependent thermal time* (Connes and Rovelli, 1994). That is, different preferred states give different time structures. In the most non-commutative[43] setting of the theory with an observable algebra $\mathfrak{M}$ of type *III* the group of modular automorphisms is unique *modulo* the inner automorphisms of the algebra, so that the thermal time is uniquely given, hence *state-independent*.

Since for the description of matter in *equilibrium* the concept of the Now is inapplicable, the pertinent correlations (57b) refer to the *B-time operator* T^+. Both the evolution operator $e^{-2\pi i\tau H/h}$ of the modular automorphism group of the material domain and the unitary operator $e^{-2\pi i\tau\Lambda^+}$ of the B-time evolution with the positive frequency operator Λ^+ allow analytic continuations into the complex strip $\{\tau - i\beta\,|\,\tau \in \mathbb{R}, \beta > 0\}$, necessary for the description of thermal equilibium at a temperature proportional to β^{-1}.

12 Concluding Remarks

In an unstructured *unus mundus*, notions like mind, matter, energy, or time have no *a priori* meaning. As Wigner (1949, p. 521) stated:

> "The world is very complicated and it is clearly impossible for the human mind to understand it completely. Man has therefore devised an artifice which permits the complicated nature of the world to be blamed on something which is called accidental and thus permits him to abstract a domain in which simple laws can be found."

---

[43] Only the factors of type I_2 and *III* are "most non-commutative" in the sense that for each non-trivial projection P there exists a maximally incompatible projection Q (see Raggio and Rieckers, 1983, p. 284).

Fundamental physics gets its simple laws by artificially postulating a *principle of uniformity of nature* (that is, invariance with respect to displacements in time and space), thereby suppressing indexical and intentional features. That is, *the laws of physics are not laws of nature, but they prompt scientists to act.*

Contemporary physics is based on the regulative idea that its basic laws do not change. This necessitates the introduction of a domain which is ordered by a correlation parameter, usually called *homogeneous time*, which we refer to as *B-time.* Of course, physical laws which are invariant under translation and reversal of B-time cannot give a comprehensive description of natural phenomena since *there is no fundamental physical principle that is related to the concept of a Now.* In experimental science the role of mental activity and intentionality cannot be neglected. For example, the first principles of physics alone do neither allow to distinguish cause and effect nor to specify initial conditions. Though the principles of physics are of great importance, they allow only an incomplete description of the world. The inevitable complement is the mental domain with a tensed A-time, required for the description of intentionality and other phenomena of consciousness.

The traditional difficulties with the concepts “A-time” and “B-time” arise because they cannot be captured within a single Boolean description. But they can be conceived in terms of a non-Boolean description generated by the affine Weyl–Heisenberg symmetry group. Epistemically accessible partial descriptions can then be generated by an epistemic breaking of the full temporal symmetry. The two affine subgroups of the affine Weyl–Heisenberg group are complementary in a mathematically well-defined sense and allow a precise description of A-time and B-time, respectively. It follows that both A-time and B-time are necessary but none of them has a privileged status, none of them can replace the other.

It is important not to forget that our distinctions between an atemporal material, an atemporal mental, and a temporal domain do not imply an ontological partition of the world – it is chosen as a partition of the universe of discourse to facilitate the discussion. Likewise the distinction between A-time and B-time refers merely to contextually meaningful *descriptions.* These descriptions are conditional on the chosen partition into material and mental domains. Thereby mind and matter appear as complementary and holistically correlated aspects of the same transcendental non-Boolean reality while time arises as an order parameter related to the breakdown of its holistic symmetry.

Probably it does not make sense to assume that the material and the mental domain are interacting, but in our description it is most natural to expect that these domains are correlated (in the sense that typically states on $\mathfrak{M} \bar{\otimes} \mathfrak{T} \bar{\otimes} \mathfrak{N}$ are not product states). Such holistic correlations are difficult to comprehend in terms of the language of the chosen description. Other divisions of the universe of discourse which lead to alternative complementary descriptions are logically possible. For a deeper understanding we may have to look for such alternative divisions, leading to different complementary viewpoints.

Acknowledgement

It is a pleasure to thank Harald Atmanspacher for reading and providing important critical remarks on earlier versions of this paper.

References

Amann, A. (1986): Observables in W*-algebraic quantum mechanics. *Fortschritte der Physik* **34**, 167–215.

Bergson, H. (1911): *Creative Evolution.* Holt, New York.

Bernays, P. (1935): Sur le platonisme dans les mathématiques. *L'enseignement mathematique* **34**, 52–69.

Bernays, P. (1946): Review of "Russell's mathematical logic" by Kurt Gödel. *Journal of Symbolic Logic* **11**, 75–79.

Bhat, B.V.R. (1996): An index theory for quantum dynamical semigroups. *Transactions of the American Mathematical Society* **348**, 561–583.

Bohr, N. (1948): On the notions of causality and complementarity. *Dialectica* **2**, 312–319.

Born, M. and Wolf, E. (1959): *Principles of Optics.* Pergamon Press, London.

Bunge, M. (1980): *The Mind–Body Problem. A Psychobiological Approach.* Pergamon Press, Oxford.

Busch, P. and Lahti, P.J. (1986): To what extent do position and momentum commute? *Physics Letters* **A115**, 259–264.

Cantor, G. (1883): Über unendliche, lineare Punktmannigfaltigkeiten, Nr. 5. *Mathematische Annalen* **21**, 545–586.

Carnap, R. (1963): Intellectual autobiography. In: Schilpp, P.A. (ed.), *The Philosophy of Rudolf Carnap.* Open Court, La Salle, Illinois, pp. 1–84.

Cohen, L. (1995): *Time–Frequency Analysis.* Prentice Hall, Upper Saddle River (New Jersey).

Connes, A. and Rovelli, C. (1994): Von Neumann algebra automorphisms and time-thermodynamics relation in generally covariant quantum theories. *Classical and Quantum Gravity* **11**, 2899–2917.

Crane, T. and Mellor, D.H (1990): There is no question of physicalism. *Mind* **99**, 185–206.

Curie, P. (1894): Sur la symétrie dans les phénomènes physiques. *Journal de Physique* **3**, 393–416.

Davies, E.B. (1976): *Quantum Theory of Open Systems.* Academic Press, London.

Dolph, C.L. (1963): Positive real resolvents and linear passive Hilbert systems. *Annales Academiae Scientiarum Fennicae, Series A. I. Mathematica* **No. 336/9**, 1–39.

Edelman, G.M. (2005): *Wider Than the Sky. A Revolutionary View of Consciousness.* Penguin Books, London.

Einstein, A. (1934): On the method of theoretical physics. *Philosophy of Science* **1**, 163–169.

Engel, K.-J. and Nagel, R. (2000): *One-Parameter Semigroups for Linear Evolution Equations.* Springer, New York.

Gabor, D. (1946): Theory of communication. *Journal of the Institution of Electrical Engineers (London)* **93, Pt.III**, 429–457.

Gale, R.M. (1968): *The Language of Time*. Routledge and Kegan Paul, London.

Gelfand, I. and Neumark, M. (1947): Unitary representations of the group of linear transformations of the straight line. *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS* **55**, 567–570.

Good, I.J (1962): Two-way determinism. In: Good, I.J. (ed.), *The Scientist Speculates*. Heinemann, London, pp. 314–315.

Gröchenig, K. (2001): *Foundations of Time–Frequency Analysis*. Birkhäuser, Boston.

Grünbaum, A. (1973): *Philosophical Problems of Space and Time*. Second enlarged edition. Reidel, Dordrecht.

Grossmann, A., Morlet, J. and Paul, T. (1986): Transforms associated to square integrable group representations. II. Examples. *Annales de l'Institut Henri Poincaré* **A 45**, 293–309.

Haag, R., Hugenholtz, N.M. and Winnink, M. (1967): On the equilibrium states in quantum statistical mechanics. *Communications in Mathematical Physics* **5**, 215–236.

Haag, R., Kastler, D. and Trych-Pohlmeyer, E.B. (1974): Stability and equilibrium states. *Communications in Mathematical Physics* **38**, 173–193.

Hartley, R.V.L. (1928): Transmission of information. *Bell System Technical Journal* **7**, 535–563.

Havin, V. and Jöricke, B. (1994): *The Uncertainty Principle in Harmonic Analysis*. Springer, Berlin.

Heil, J. (1998): *Philosophy of Mind: A Contemporary Introduction*. Routledge, London.

Heisenberg, W. (1959): Wolfgang Paulis philosophische Auffassungen. *Naturwissenschaften* **46**, 661–663.

Hempel, C.G. (1980): Comments on Goodman's "Ways of Worldmaking". *Synthese* **45**, 193–199.

Hlawatsch, F., Tauböck, G. and Twaroch, T. (2003): Covariant time–frequency analsyis. In: Debnath, L. (ed.), *Wavelets and Signal Processing*. Birkhäuser, Boston, pp. 177–231.

Hogan, J.A. and Lakey, J.D. (2005): *Time–Frequency and Time–Scale Methods*. Birkhäuser, Boston.

Howe, R. (1980): On the role of the Heisenberg group in harmonic analysis. *Bulletin of the American Mathematical Society* **3**, 821–843.

Howland, J.S. (1974): Stationary scattering theory for time-dependent Hamiltonians. *Mathematische Annalen* **207**, 315–335.

Husserl, E. (1966): *Zur Phänomenologie des inneren Zeitbewusstseins (1893–1917)*. Herausgegeben von Rudolf Boehm. Martinus Nijhoff, Haag.

Husserl, E. (1983): *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. First Book*. Translated by F. Kersten. Kluwer, Dordrecht.

Inwagen, P. van (1975): The incompatibility of free will and determinism. *Philosophical Studies* **25**, 185–199.

James, W. (1890): *The Principles of Psychology. Volume One*. Holt, New York.

Johnson, H.J. (1973): Changing concepts of matter from antiquity to Newton. In: Wiener, P.P. (ed.), *Dictionary of the History of Ideas, Vol.III*. Charles Scribner's Sons, New York, pp. 185–196.

Jung, C.G. (1970): *The Collected Works of C.G. Jung. Volume 14. Mysterium Conjunctionis.* Second enlarged edition. Princeton University Press, Princeton.

König, H. and Tobergte, J. (1963): Reversibilität und Irreversibilität von linearen dissipativen Systemen. *Journal für die reine und angewandte Mathematik* **212**, 104–108.

Kadison, R.V. and Ringrose, J.R. (1986): *Fundamentals of the Theory of Operator Algebras. II. Advanced Theory.* Academic Press, Orlando.

Kalisa, C. and Torrésani, B. (1991): N-dimensional affine Weyl–Heisenberg wavelets. *Annales de l'Institut Henri Poincaré* **A 59**, 201–236.

Kanwal, R.P. (1983): *Generalized Functions.* Academic Press, New York.

Kim, J. (1998): *Mind in a Physical World.* MIT Press, Cambridge, Mass.

Landau, H. J. and Pollak, H. O. (1961): Prolate spherical wave functions, Fourier analysis and uncertainty. II. *Bell System Technical Journal* **40**, 65–84.

Landau, H. J. and Pollak, H. O. (1962): Prolate spherical wave functions, Fourier analysis and uncertainty. III: The dimension of the space of essentially time- and band-limited signals. *Bell System Technical Journal* **41**, 1295–1336.

Lax, P.D. and Phillips, R.S. (1967): *Scattering Theory.* Academic Press, New York.

Lax, P.D. and Phillips, R.S. (1978): The time delay operator and a related trace formula. In: Gohberg, I. and Kac, M. (eds.), *Topics in Functional Analysis. Advances in Mathematics, Supplementary Studies, Volume 3.* Academic Press, New York, pp. 197–215.

Lindner, F., Schätzel, M.G., Walther, H., Baltuška, A., Goulielmakis, E., Krausz, F., Milošević, D.B., Bauer, D., Becker, W. and Paulus, G.G. (2005): Attosecond double-slit experiment. *Physical Review Letters* **95**, 040401-1–040401-4.

Lindquist, A. and Picci, G. (1991): A geometric approach to modelling and estimation of linear stochastic systems. *Journal of Mathematical Systems, Estimation and Control* **1**, 241–333.

McTaggart, J.E. (1908): The unreality of time. *Mind* **17**, 456–473.

Mellor, D.H. (1998): *Real Time II.* Routledge, London.

Montero, B. (2003): Varieties of causal closure. In: Walter, S. and Heckmann, H.-D. (eds.), *Physicalism and Mental Causation.* Imprint Academic, Exeter, pp. 173–187.

Newton, I. (1730): *Opticks*, London, fourth edition. Reprinted by Dover Publications, New York, 1952.

Nickel, G. (1996): On evolution semigroups and nonautonomous Cauchy problems. *Dissertation Summaries in Mathematics* **1**, 195–202.

Noether, E. (1918): Invariante Variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften, Göttingen. Mathematisch Physikalische Klasse* **1918**, 235–257.

Oaklander, L.N. (2004): *The Ontology of Time.* Prometheus Books, Amherst.

Oaklander, L.N. and Smith, Q. (1994): *The New Theory of Time.* Yale University Press, New Haven.

Papandreou-Suppappola, A. (2003): Time-frequency processing of time-varying signals with nonlinear group delay. In: Debnath, L. (ed.), *Wavelets and Signal Processing.* Birkhäuser, Boston, pp. 311–359.

Park, D. (1972): The myth of the passage of time. In: Fraser, J.T., Haber, F.C. and Müller, G.H. (eds.), *The Study of Time.* Springer, Berlin, pp. 110–121.

Pauli, W. (1948): Editorial. *Dialectica* **2**, 307–311.

Pauli, W. (1952): Der Einfluss archetypischer Vorstellungen auf die Bildung naturwissenschaftlicher Theorien bei Kepler. In: Jung, C.G. and Pauli, W. (eds.), *Naturerklärung und Psyche*. Rascher Verlag, Zürich, pp. 109–194.

Pauli, W. (1954): Matter. In: *Man's Right to Knowledge*. Columbia University Press, New York, pp. 10–18.

Piron, C. (1964): Axiomatique quantique. *Helvetica Physica Acta* **37**, 439–468.

Poincaré, H. (1907): *The Value of Science*. Science Press, New York.

Primas, H. (2007): Non-Boolean descriptions for mind–matter problems. *Mind and Matter* **5**, 7–44.

Prior, A. (1967): *Past, Present and Future*. Clarendon Press, Oxford.

Putnam, H. (1992): *Renewing Philosophy*. Harvard University Press, Cambridge (Massachusetts).

Raggio, G.A. and Rieckers, A. (1983): Coherence and incompatibility in W*-algebraic quantum theory. *International Journal of Theoretical Physics* **22**, 267–291.

Reed, M. and Simon, B. (1975): *Methods of Modern Mathematical Physics. II: Fourier Analysis, Self-Adjointness*. Academic Press, New York.

Reichenbach, H. (1956): *The Direction of Time*. University of California Press, Berkeley.

Rovelli, C. (1993): Statistical mechanics of gravity and the thermodynamical origin of time. *Classical and Quantum Gravity* **10**, 1549–1566.

Rovelli, C. (1995): Analysis of the distinct meanings of the notion of "time" in different physical theories. *Il Nuovo Cimento* **110B**, 81–93.

Schempp, W. (1986): *Harmonic Analysis on the Heisenberg Nilpotent Lie Group*. Longman, Harlow.

Schmüdgen, K. (1983): On the Heisenberg commutation relation. I. *Journal of Functional Analysis* **50**, 8–49.

Scully, M.O., Englert, B. and Walther, H. (1991): Quantum optical test of complementarity. *Nature* **351**, 111–116.

Sherover, C.N. (1975): *The Human Experience of Time. The Development of Its Philosophical Meaning*. Northwestern University Press, Evanston.

Slepian, D. (1976): On bandwidth. *Proceedings of the IEEE* **64**, 292–300.

Slepian, D. (1983): Some comments on Fourier analysis, uncertainty and modeling. *SIAM Review* **25**, 379–393.

Slepian, D. and Pollak, H. O. (1961): Prolate spherical wave functions, Fourier analysis and uncertainty. I. *Bell System Technical Journal* **40**, 43–64.

Synge, J.L. (1960): Classical dynamics. In: Flügge, S. (ed.), *Principles of Classical Mechanics and Field Theory. Encylopedia of Physics. Volume III/1*. Springer, Berlin, pp. 1–225.

Szriftgiser, P., Guéry-Odelin, D., Arndt, M. and Dalibard, J. (1996): Atomic wave diffraction and interference using temporal slits. *Physical Review Letters* **77**, 4–7.

Takesaki, M. (1979): *Theory of Operator Algebras I*. Springer, New York.

Tieszen, R. (2005): *Phenomenology, Logic, and the Philosophy of Mathematics*. Cambridge University Press, Cambridge.

Torrésani, B. (1991): Wavelets associated with representations of the affine Weyl–Heisenberg group. *Journal of Mathematical Physics* **32**, 1273–1279.

Vilenkin, N.J. (1968): *Special Functions and the Theory of Group Representations*. American Mathematical Society, Providence.

Wang, H. (1995): Time in philosophy and in physics: From Kant and Einstein to Gödel. *Synthese* **102**, 215–234.

Weyl, H. (1918): *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis.* Veit, Leipzig.

Weyl, H. (1924): Was ist Materie? *Naturwissenschaften* **12**, 585–568, 604–611.

Weyl, H. (1925): Die heutige Erkenntnislage in der Mathematik. *Symposion* **1**, 1–32.

Weyl, H. (1949): *Philosophy of Mathematics and Natural Science.* Princeton University Press, Princeton.

Weyl, H. (1952): *Symmetry.* Princeton University Press, Princeton.

Wiener, N. (1948): Time communication and the nervous system. *Annals of the New York Academy of Sciences* **50**, 197–220.

Wiener, P.P. (1951): *Leibniz Selections.* Charles Scribners's Sons, New York.

Wigner, E.P. (1949): Invariance in physical theory. *Proceedings of the American Philosophical Society* **93**, 521–526.

Williams, D.C. (1951): The myth of passage. *The Journal of Philosophy* **48**, 457–472.

Wollenhaupt, M., Assion, A., Liese, D., Sarpe-Tudoran, C., Baumert, T., Zamith, S., Bouchene, M.A., Girard, B., Flettner, A., Weichmann, U. and Gerber, G. (2002): Interferences of ultrashort free electron wave packets. *Physical Review Letters* **89**, 173001-1–173001-4.

A Proposed Relation Between Intensity of Presence and Duration of Nowness

G. Franck[1] and H. Atmanspacher[2]

[1] Department of Computer Aided Architecture and Planning, Vienna University of Technology, Austria, `franck@iemar.tuwien.ac.at`
[2] Institute for Frontier Areas of Psychology, Freiburg, Germany, `haa@igpp.de`

Summary. It is proposed to translate the mind-matter distinction into terms of mental and physical time. In the spirit of this idea, we hypothesize a relation between the intensity of mental presence and a crucial time scale (some seconds) often referred to as a measure for the duration of nowness. This duration is experimentally accessible and might, thus, offer a suitable way to characterize the intensity of mental presence. Interesting consequences with respect to the idea of a generalized notion of mental presence, with human consciousness as a special case, are outlined. Our approach includes some features consistent with other, related ideas which are indicated.

1 Introduction

In recent years, the mind-matter distinction is often referred to in terms of *dual-aspect* or *dual-perspective* accounts. In such frameworks, the mental and the material are assumed to be aspects of an underlying non-dual entity whose ontological nature has so far resisted a unifying descriptive account. A particularly fashionable dual-perspective account refers to first-person and third-person perspectives. In the perspective of the first person, subjective experience and mental phenomena are accessible. The perspective of the third person, by virtue of intersubjective operationalization, can be utilized for a specific way to address the material domain.

If we restrict ourselves to the discussion of mind and brain, the neural, chemical, and physical processes going on in the brain are assumed to belong to the material domain, whereas subjective experiences, also denoted as *qualia*, are appearances in *mental presence*. Insofar as qualia have no referents in material reality, they are "epistemically empty" concerning this reality, in the same sense as illusions or dreams are. However, they carry another kind of epistemic content called *phenomenal content*.

As an alternative to dual-aspect or dual-perspective accounts, it has occasionally been proposed to translate the mind-matter distinction into terms of time (for more recent accounts cf. Franck, 2004, 2008; Primas, 2003, 2008): Mental presence is addressed in terms of mental time while material reality is addressed by physical time. We consider this proposal as particularly promising because time plays a substantial role in *both* the mental *and* the material domain, yet this role shows characteristic differences in the two domains. The discussion of mental time and physical time has been a central and controversial topic in the philosophy of science since Mach, Russell, Einstein, and McTaggart (cf. Reichenbach, 2000; Grünbaum, 1963; Whithrow, 1980; Denbigh, 1981; Jammer, 2006).

2 Tensed Time and Tenseless Time

There are two important ways to address and compare key features of mental and physical time. One of them, originating in the philosophy of language, starts with the notion of tense. Briefly speaking, tensed time is a notion of mental time exhibiting the regions of *past* and *future* separated by the *present now*. By contrast, physical time is tenseless and is limited to the relations of "earlier than", "later than", and "simultaneous with". The other starting point lies in physics, where a number of symmetry or invariance principles can be used to characterize features of time. Such principles express what remains unchanged if particular parameters are varied. Most significant examples are time-translation invariance, time-reversal invariance, and time-scale invariance. Let us begin with a discussion of these invariances for mental and physical time in more detail.

It is well known that all fundamental laws in physical theories are *time-translation invariant*, i.e., they are independent of a particular instant in time t_{initial} serving as an initial condition for their solution. This implies that the fundamental laws are independent of the time at which their predictions are empirically investigated. In this sense the choice of t_{initial} is arbitrary, and there is no present or nowness in fundamental physical theories.

Moreover, the fundamental laws of physics are also *time-reversal invariant*. This is to say that, for any arbitrarily chosen instant $t_{\mathrm{initial}} = 0$, their solution in one direction of time has a time-reversed copy which is equally feasible. This feature is at variance with the empirical observation of a distinguished forward direction of processes in time from $t < 0$ to $t > 0$. (Note that the notions of past and future are illegitimate in tenseles physical time.) This directedness, often called *irreversibility*, is standardly explained by particular initial and boundary conditions.

Both time-translation invariance and time-reversal invariance indicate an important difference between theoretical and experimental physics. Since every experiment is carried out at a particular date and with non-anticipative measuring instruments, both invariances are broken in experimental physics (cf. Primas, 2008). In contrast to the fundamental laws of physics, experimental physics contains, thus, the notions of nowness and irreversibility (of course without any phenomenal content as it occurs in the subjective experience of qualia).

These two notions are even more crucial, if the focus is shifted from physical tenseless time to mental tensed time. From the point of view of tensed time, including the subjective experience of temporal sequences of mental states, there are two basically unquestioned features: (i) Any stage during such a sequence refers to a *present now* that distinguishes past and future; (ii) Any such sequence is *irreversibly directed* from past to future. In addition to nowness and irreversibility in experimental physics, their *mental* significance contains the qualitative character of a subjective experience, its phenomenal content or its quale. This aspect is deliberately disregarded in any description of physical or otherwise material systems, including the brain.

Scale invariances play a role wherever there is no intrinsically prescribed unit of measure, i.e. no intrinsic length or time scale. For instance, the unit of a second in physical time measurement is arbitrary in the sense that physical processes are not organized in such a way that a second would be a distinguished measure of time. The recent literature on self-affine or self-similar structures provides a bunch of illustrative examples for scale-invariant phenomena. Time-scale invariance together with time-translation invariance constitutes a group of transformations which is called an affine group. It means that displacements in time and stretching or squeezing time intervals makes no difference for the description of the considered process.

3 Can the Intensity of Presence Be Measured ?

So far we have described how a translation of the mind-matter distinction to the distinction of mental tensed time and physical tenseless time leads to characteristic though subtle differences between the two notions of time. Now we want to (i) identify general qualitative features of mental presence that can be related to properties of tensed time and (ii) look for options to express these properties in terms of tenseless physical time in order to operationalize them.

Mental presence is at the basis of all subjective experience, manifesting itself in a variety of possible ways. In this sense, presence can be conceived as a most fundamental quale, within which the appearance of more specific qualia becomes possible. The concept of mental presence as such does not necessarily imply an explicit experience *of* something, but should be understood as an immediate and implicit "being aware". This awareness does not require a self-model, let alone an explicit representation of such a self-model (self-consciousness). It may be as primitive as the "creature consciousness" that Chalmers (2000) suggests as the most primitive form of conscious experience conceivable.[3] Creature consciousness amounts to nothing but an awareness without any differentiation concerning a self that is aware and an intentional content that it is aware of. In creature consciousness, the awareness of presence and the presence of awareness coincide. Nevertheless, the presence of an awareness, however primitive and dim it may be, amounts to an experience of "how it is like to be" an experiencing subject.

Chalmers (2000) locates creature consciousness at the lowest level in a hierarchy of what he calls *phenomenal families*. The next-to-lowest level considers the distinction of sleep and wake states of consciousness, which are typically subject to circadian cycles. Within wake states one can, e.g., distinguish motivational, emotional, and cognitive states, and within those one can move to more and more specific representations with qualia. This way, a hierarchy of phenomenal contents with increasing differentiation emerges. The state of current research does not provide much concrete material to assess the levels of phenomenal families in a consistent and detailed way. However, it seems plausible to assume that the awareness of presence, which is associated with a particular phenomenal family, becomes more intense when moving up the hierarchy.[4]

Beyond these differentiations, which may be referred to as "phenomenal changes", there are two additional possibilities to grade the intensity of presence. The first one is due to the amount of *attention* with which a state of consciouness is focused at.[5] Corresponding variations of the intensity of presence are called "focal changes" and can be accomplished in a more or less controlled (voluntary) fashion, depending on the degree of vigilance. The second kind of gradation is due to the distance of a considered phenomenal content from the temporal present. Clearly, the intensity of presence is highest if a quale is just experienced, i.e. located in the now, and the intensity decreases with growing distance from the now (memory toward past, anticipation toward future). Corresponding variations of the intensity of presence are called "temporal change". They occur autonomously because the now moves independently of a subject's attentional control.

In this way, *we have identified an important interface between mental presence and temporal present, with attentional focus as a potentially moderating factor*. If the intensity of presence can indeed be related to nowness and attention, the next step is to think about possible ways how this can be fleshed out. First of all, this means that we have to think about ways in which nowness and attention can be evaluated quantitatively or at least quasi-quantitatively. If such evaluations turn out to be possible, they provide interesting candidates to study the intensity of presence, even though indirectly.

---

[3] Similarly, Edelmann and Tononi (2000) speak of "primary consciousness", and Damasio (1999) speaks of "core consciousness".

[4] The notion of an *intensity of presence* does, of course, need to be defined more precisely. For more discussion see Franck (2008), and for some additional details see Metzinger (2003), pp. 184–189.

[5] A comprehensive account of the psychology of attention is due to Pashler (1998).

In the following section, we popose a way in which quantitative measures for the duration of nowness might be related to the degree of attention (Atmanspacher *et al.*, 2004, 2008a). The model is called Necker-Zeno model and was originally designed to describe the dynamics of the bistable perception of ambiguous stimuli. The model is formulated exclusively in terms of physical, tenseless time. We will therefore have to argue that the variables of the model can be related to tensed time and mental presence, thus approaching an empirical operationalization of quantifiable aspects of qualitative mental concepts.

4 Time Scales in the Necker-Zeno Model

4.1 Review of the Necker-Zeno Model

The Necker-Zeno model (Atmanspacher *et al.*, 2004, 2008a,b) is inspired by the quantum Zeno effect (Misra and Sudarshan, 1977) and describes the bistable perception of ambiguous stimuli such as the Necker cube (Necker 1832) in a formal fashion. In contrast to attempts to apply standard quantum physics to brain functioning and consciousness directly, the Necker-Zeno model is based on a generalized formal framework, particularly suited for applications *beyond* physics (Atmanspacher *et al.*, 2002). Earlier suggestions to use Zeno-type arguments for cognitive systems are due to Ruhnau (1995) and Stapp (1999).

A key assumption of the Necker-Zeno model is that the cognitive state corresponding to a perceived stimulus is updated at intervals ΔT (of the order of 30 msec to 70 msec, see below). The probability that no reversal occurs within a time period T is then given by:

$$w(T) \;=\; \cos^2(gT) \quad \text{with} \quad g \;=\; \frac{\pi}{4t_0}, \tag{1}$$

where t_0 characterizes the period of the reversal dynamics without updates (of the order of 300 msec, see below). The inverse of t_0, g, determines how fast the cognitive state corresponding to a perceived stimulus decays.

Let $\{\tau_i\}_{i=0,...,N}$ be the instants at which an update of the cognitive state has been performed, and let $w(\tau_N, \tau_{N-1}, ..., \tau_1, \tau_0 = 0)$ be the joint probability that no perceptual reversal has occured from τ_0 up to $\tau_N = T$. Then

$$W(T) := w(\tau_N, \tau_{N-1}, ..., \tau_1) = \prod_{i=1}^{N} \cos^2\{g(\tau_i - \tau_{i-1})\} = \prod_{i=1}^{N} \cos^2\{g\Delta T(i)\} \;\;,$$

with

$$\Delta T(i) = \tau_i - \tau_{i-1}.$$

For the Necker-Zeno model we have $\Delta T(i) \ll t_0$, so we may expand the cosine up to the quadratic term:

$$W(T) \;\approx\; e^{2\ln(1-\frac{1}{2}g^2(\Delta T_i)^2)} \;\approx\; e^{-g^2\sum_{i=1}^{N}(\Delta T_i)^2}.$$

Assuming a constant updating interval $\Delta T(i) = \Delta T$, we obtain

$$W(T) \;=\; e^{-g^2 N(\Delta T)^2},$$

which means for $T = N\Delta T$:

$$W(T) \;=\; e^{-g^2 \Delta T \cdot T}. \tag{2}$$

$W(T)$ is the probability that no reversal has occurred up to time T. Hence, $1 - W(T)$ describes the integrated (cumulative) distribution of "dwell times" (inverse reversal rates). It yields the following probability distribution (density) for dwell times:

$$P(T) \;=\; -\frac{\mathrm{d}W(T)}{\mathrm{d}T} \;=\; \gamma\, e^{-\gamma T}, \tag{3}$$

where $\gamma = g^2\Delta T$. The mean dwell time $\langle T\rangle$ is given by:

$$\langle T\rangle \;=\; \frac{1}{\gamma} \;=\; \left(\frac{16}{\pi^2}\right)\frac{t_0^2}{\Delta T}, \tag{4}$$

leading to the relation

$$\Delta T \cdot \langle T\rangle \;=\; C\, t_0^2, \tag{5}$$

where C is of the order of 1 such that t_0 is basically the geometric mean of $\langle T\rangle$ and ΔT.

In this way, the Necker-Zeno model predicts a quantitative relationship between three time scales which can be interpreted in terms of cognitive time scales (for more details see Atmanspacher *et al.*, 2004):

(i) The time between successive information updates of the cognitive state is related to the so-called *sequential order threshold* of $\Delta T \approx 30$ msec (Pöppel 1997). In the original quantum Zeno effect ΔT is the time between successive observations.

(ii) The *decay time* for a sensory input to become consciously accessible (cognitively processed) is of the order of $t_0 \approx 300$ msec (Basar-Eroglu *et al.*, 1993). In the original quantum Zeno effect t_0 is the oscillation period between the two unstable states without updates, a situation which is of more or less hypothetical character in cognition.

(iii) The observed *mean dwell time* $\langle T\rangle$ between successive reversals of competing configurations of an ambiguous stimulus is usually of the order of 3 sec (Pöppel 1997). $\langle T\rangle$ has often been referred to as a duration of an "extended nowness" that is not restricted to a point separating past and future but covers a temporal interval.

These cognitive time scales obviously satisfy Eq. (5). More detailed empirical tests of Eq. (5) are possible if one of the time scales can be measured as a function of another one, which is experimentally controllable, while the third one is considered fixed.

In this respect, a number of model predictions have been confirmed by results from experiments carried out under discontinuous stimulus presentations. Atmanspacher *et al.* (2004) showed that Eq. (5) describes the behavior of $\langle T \rangle$ as a function of interstimulus intervals (off-times) $t_{\rm off}$ greater than t_0. More recently, it has been shown Atmanspacher *et al.* (2008a) that the model describes the behavior of $\langle T \rangle$ as a function of $t_{\rm off} < t_0$ as well. These results are non-trivial since they represent opposing trends for long and short off-times, separated by a critical time scale of the order of 300 msec. Atmanspacher *et al.* (2008a) also demonstrated that the empirically observed distribution of dwell times $P(T)$ or, respectively, inverse reversal rates, is matched by the model. This can be achieved by considering an initial (transient) phase for the reversal dynamics, which is highly plausible. The initial behavior can be implemented in terms of an initial decrease of ΔT or an initial increase of t_0 up to a time after which their asymptotic values are reached.

For a cognitive interpretation of the initial phase of the reversal dynamics, Atmanspacher *et al.* (2008a) speculated that some kind of attention relaxation may be a significant factor. For instance, recent evidence (van Ee 2005, Meng and Tong 2004) for voluntary control over dwell times in the perception of ambiguous figures – as opposed to binocular rivalry – could imply a significant contribution of top-down processing – as opposed to bottom-up processing. The time scales involved should, thus, be longer for bistability in ambiguous perception. Moreover, Reisberg and O'Shaughnessy (1984) found that dwell times increase if attention is distracted, and Vickers (1972) found that dwell times are reduced by increasing vigilance.

4.2 Duration of Nowness and Degree of Attention

The elementary quale of an intensity of presence can vary in three different but not independent ways briefly introduced in Sec. 3: (a) due to a change in distance from the temporal present, (b) due to a change of focal attention, and (c) due to a change of the phenomenal family. The attempt now is to make the step from these mental characterizations to physical features that are capable of operationalizing changes of the intensity of presence in the mentioned varieties.

We propose that an interesting candidate for this purpose is the duration of nowness, measurable in the sense of a physical clock time. Note that this attempt deliberately disregards the phenomenal (qualia) features of the subjective experience of nowness and focuses on an accessible physical correlate. Thus, it does in principle not differ from disregarding phenomenal (qualia) features of the experience of color such that physical characterizations like wavelengths remain. (We argue, however, that the quale of nowness should be considered much more fundamental than, e.g., qualia of color.)

a) A change in distance from the temporal present, briefly temporal change, is simply caused by the fact that the temporal present, briefly the now, moves along with physical time. We do not discuss the question why and how the moving now is synchronized with physical time (see Primas, 2003, who focuses mainly on temporal change). The intensity of presence for a perceived phenomenon increases while approching the center of the window of nowness and decreases subsequently until it fades away when the nowness interval $\langle T \rangle$ is left.[6]

The experience that something can change at all (as a function of physical time) depends fundamentally on the experience of temporal change that is associated with the change of the elementary quale of the intensity of presence. Since temporal change is autonomous, it cannot be influenced voluntarily and, thus, cannot be used as an experimentally controllable independent variable. Although a variable size of $\langle T \rangle$ expresses that nowness intervals can have variable extension, the cause of this variability remains unclear.

b) A change of focal attention, briefly focal change, occurs always in front of the background of autonomous temporal change. But different from temporal change, the intensity of presence is now susceptible to control by attention. The Necker-Zeno model provides two options to influence the duration $\langle T \rangle$ of nowness (Eq. (5)): (1) A decreasing updating interval ΔT, corresponding to a high level of attention, leads to increasing $\langle T \rangle$ if t_0 is constant; (2) An increasing decay time t_0, also corresponding to high attention, leads also to increasing $\langle T \rangle$ if ΔT is constant.[7]

Both options, changing t_0 or changing ΔT, offer an experimentally feasible operationalization of the intensity of presence (via focal change) in terms of the duration of nowness. Focal change reduces to temporal change as voluntary attention ceases. The difference between the two kinds of change has an important subjective aspect: the experience of agency in focal change, which is lacking in temporal change. However, the capability of voluntary control also depends on the degree of vigilance: the intensity of presence in Chalmers' (2000) background states of consciousness.

Roughly speaking, there is a continuous spectrum of degrees of vigilance between wake states and sleep states within the phenomenal family of background consciousness. Falling asleep means to loose control over the focus of attention. Nevertheless, the experience of temporal change can continue,

---

[6] Husserl (1996) denoted these two stages as protention and retention. Varela (1999) gives a detailed account of how Husserl's concepts can be related to cognitive neuroscience and indicates the perception of multistable stimuli as a promising field in this respect.

[7] Recent results by Carter *et al.* (2005) show that $\langle T \rangle$ can in fact be enormously extended (by a factor of 1000) as compared to normal conditions. Atmanspacher *et al.* (2008a) indicated empirical evidence that an attention-driven change of t_0 is more likely to affect $\langle T \rangle$ than a change of ΔT.

for instance in dreams. In dreamless sleep, this experience is extinguished. Dreamless sleep is the lowest level that vigilance reaches in a circadian cycle. We do not know yet whether this results in an expansion or a contraction of the duration of nowness, but in principle this question is open to empirical research.

c) A change of the phenomenal family to which the experienced phenomena belong is the third option in which the intensity of presence can vary. Moving up the hierarchy of phenomenal families can to some extent be associated with a varying degree of vigilance as a measure of the intensity of presence. At higher phenomenal levels, the intensity of presence consists of both a (usually) high degree of vigilance and the intensity with which a phenomenal content is present.

Even though these issues are purely speculative at present, it is plausible to assume that the duration of nowness $\langle T \rangle$ at a low-level phenomenal family, close to what Chalmers calls creature consciousness, is smaller than at higher levels, where substantial differentiations of phenomenal content abound and require higher degrees of attention and more intense mental presence. This leads to the question at which level of "creatures" one is entitled to speak about "creature consciousness" at all. Could there be a level of "proto-mental" presence below the level of creature consciousness ? Could such a "proto-mental" presence be the fundamental and ubiquitous feature of the universe from which creature consciousness emerges at specific degrees of complexity in the organization of matter ?

4.3 Operationalizing Panpsychism ?

The doctrine that mind is a fundamental feature of the world, which exists throughout the universe, is called panpsychism (Seager and Allen-Hermanson 2001, Skrbina 2005). In a recent paper, Strawson (2006) made a comprehensive attempt to defend panpsychism, for instance he says that "everything that concretely exists is intrinsically experience-involving" (p.8). We do not intend to go into Strawson's arguments in detail here, but the quoted strong statement can easily be interpreted in a way that is very similar to the approach offered in this paper. The similarity is most clearly visible if Strawson's notion of "concreteness" is understood as "being in presence". Our notion of an intensity of presence would then be equivalent to a "degree of concreteness" in Strawson's approach.[8]

As an immediate consequence, Strawson's notion of concreteness would have to be differentiated according to temporal change, focal change, and change of the phenomenal family involved. With particular respect to the last, the question addressed at the end of the preceding subsection becomes crucial. At which level of description should we assume that creature consciousness

---

[8] Note that Strawson (2006) does not try to specify his notion of concreteness or even discuss its potential gradation.

enters? There are basically two possibilities to answer this question. One of them, the standard position of panpsychism, is that some rudimentary form of mental activity is a fundamental and ubiquitous feature of the universe. In other words, something "protomental" or "protoexperiential" is engrained in *every* element of material reality.

The other possibility is that the emergence of creature consciousness in the course of biological evolution required some critical degree of complexity in the organization of matter. No specification of such a critical degree has been convincingly demonstrated so far. Moreover, defenders of standard panpsychism maintain that experience as a fundamental feature of the world is qualitatively so different from matter that it would simply be a category mistake to consider it as an emergent feature.

The key issue of full-blown panpsychism is the question of "how it is like to be" a worm, an amoeba, a cell, or a molecule. Our proposal in this regard is, first of all, to rephrase notions such as protomentality or protoexperience in terms of extremely low degrees of an intensity of presence.[9] Then, in the spirit of the arguments given above, we suggest to operationalize the intensity of presence in temporal terms. More precisely, we suggest to apply the duration of nowness $\langle T \rangle$ in order to parametrize the full spectrum between the more developed conscious experience of mammals and the much less developed "protoexperience" of simpler organisms or elements of material reality.

In this way, we avoid the necessity of responding to the metaphysical issue of where mentality or protomentality ends or starts, and replace it by the criterion of a potentially measurable size of the duration of nowness. As a consequence, we would have $\langle T \rangle \to \infty$ for the limit of a (not-yet-observed) most developed form of conscious experience, and $\langle T \rangle \to 0$ for the limit of vanishing protoexperience. It should be emphasized again, that the latter case of an extension-free now is still outside the domain of physical theory, where there is no place for *experienced* nowness (and tense) at all. Nevertheless, our proposal suggests a smooth transition to physical theory insofar as the duration of nowness as a physical correlate of the intensity of mental presence would have the appropriate limit.

5 Relations to Other Approaches

5.1 Quantum Process Ontology

Process ontology basically argues that the fundamental elements of reality are to be conceived in terms of process rather than substance. Using corresponding ideas of James and Whitehead, Stapp (2007) has developed a comprehensive framework that (1) relates this idea to quantum theory and (2)

---

[9] This should be compared with other ideas of how to identify hallmarks of human, mammalian, and non-mammalian "consciousness", reviewed by Beshkar (2008). See also Seth *et al.* (2005), Edelman *et al.* (2005) for the same topic.

allows a discussion of mind-matter issues. One of Stapp's key assumptions is that the conventional formalism of quantum theory (*à la* von Neumann 1932) does not need to be changed for this purpose. What he advocates, however, is (a) an ontological foundation of the standard epistemological interpretation of quantum theory by Whitehead's process ontology, and (b) an addition of psychological features pertaining to the mental domain due to James.

The fundamental elements of reality that Stapp adopts from Whitehead are called actual occasions (cf. Klose, 2008). They are endowed with mental and physical poles, thus referring to mental and material aspects of reality, or consciousness and brain, respectively, in a narrower perspective. Insofar as every actual occasion has both poles, Whitehead's ontology is a paradigm example of panpsychism. Stapp deviates from this radical version: He argues that there should be a limit below which it is not reasonable to speak of mentality, or protomentality.

Another key feature of actual occasions is that they have spatial and temporal extension. The latter, which is related to James's notion of a specious present, reflects the idea that tensed time contains a temporal present that is not conceived as a point between past and future. It has a finite duration which depends on the actual occasion concerned. Whitehead does not indicate details concerning the concrete factors that may determine the duration of the present.

A specific feature of Stapp's approach, however, can be interpreted in that way. He supposes that intrinsically unstable quantum states of neuronal assemblies (involving some 10^3 to 10^6 neurons that are functionally coupled) are stabilized by the quantum Zeno effect. The strength of this effect, on Stapp's account, is related to the attentional effort with which the mental correlate of the considered neuronal assembly is focused at. Although Stapp does not explicitly refer to the duration of nowness in this context, such an interpretation may be legitimate. It would indicate that an increased degree of attention corresponds to a prolonged duration of nowness.

This agrees with the predictions, outlined in Sec. 4, according to the Necker-Zeno model. There is, however, a significant difference between this model and Stapp's implementation of the quantum Zeno effect acting on neuronal assemblies. While Stapp refers to the Zeno effect in the sense of conventional quantum theory, the Necker-Zeno model is embedded within a generalized quantum theory (Atmanspacher *et al.*, 2002), designed to address situations *outside* conventional quantum theory in particular. The example of bistable perception as a cognitive phenomenon has been worked out independently of conventional quantum theory. The Necker-Zeno model for bistable perception provides a system-theoretic description and does not assume quantum states of the brain or parts of it.

5.2 Group Representations of Tensed Time Observables

An entirely different approach discussing tense and nowness in relation to tenseless time is due to Primas (2003, 2008). He describes tensed time in terms of a Kolmogorov structure, representing an abstract type of mental memory which defines sequential order via the growth of the set of mental events. This Kolmogorov structure is associated with a non-commutative *time observable* inducing a tenseless time variable with a distribution that has non-vanishing dispersion. The spectrum of the time variable degenerates into a dispersion-free parameter in the classical (commutative) limit.

Since the tensed time observable is not commutative, tensed time can be entangled with the time variable of the material domain even if the tensed and the tenseless system do not interact. Due to this time-entanglement, the dynamical aspects of conventional quantum physics can be described in terms of strict correlations between the tensed system and the tenseless system. In the limit of vanishing correlations, the usual equations of motion of physics are recovered with an emergent parameter time as independent variable. Time-entanglement provides a reason why mental time and physical time are synchronized.

On Primas's account, tensed time T together with its complementary observable, a frequency Λ, and a scaling parameter S are proposed to generate an affine group. Primas proposes to understand the subgroups generated by T, S and Λ, S as referring to tensed and tenseless time, respectively. For the subgroup characterizing tensed time a self-adjoint Λ is not defined and, vice versa, for the subgroup characterizing tenseless time a self-adjoint T is not defined. These features eventually express the complementarity of the mental and the physical.

The tensed time subgroup has three inequivalent irreducible representations on $\mathbb{R}$ (Gelfand and Neumark 1947) which can be understood to distinguish the cases $T = 0$, $T > 0$, and $T < 0$. This provides a natural way to break the time-translation invariance, $t \rightarrow t + \tau$, of the tensed time group by introducing a temporal present, and to break the time-reversal invariance, $t \rightarrow -t$, of the tensed time group by defining future and past.

Of particular interest in our context is the "trivial" representation $T = 0$ corresponding to the present. If the spectrum of T is not dispersion-free, $T = 0$ corresponds to an extended now. The time-scale invariance of the tensed time group can be interpreted as an invariance under scaling, $T \rightarrow \sigma T$, of the extension of the now. The motion of the now (along with physical time) is subject to a broken time-reversal invariance, i.e. is directed from past to future.

Since Primas' outline is designed on a fundamental level of description, where details of concrete systems are disregarded, the parameters in his approach remain unspecified. While the parameter τ is assumed to move along with physical time, we propose that σ could be phenomenologically determined by possible operationalizations of the intensity of presence. This amounts to fixing a time scale and breaks time-scale invariance.

In Secs. 3 and 4 we discussed variations of attentional focus as a promising candidate to implement this idea empirically. In the framework of the Necker-Zeno model, we predict that such variations lead to measurable changes of the duration of nowness. Experimental studies in this direction will open a new road to studying particular aspects of mind-matter relations.

5.3 Relaxation Processes for Neural Time Keeping

In addition to the significance of nowness in mental activity, irreversibility is a key feature of subjective experience. An approach stressing the breakdown of time-reversal invariance for the neural correlates of such experiences is due to recent work summarized by Wackermann and Ehm (2006). A quick look at the mathematical representation of dynamical laws in terms of exponential functions,

$$f(t) = e^{(i\omega - \alpha)t},$$

where $\omega > 0$ is a frequency and $\alpha > 0$ is a damping rate, reveals basically two elementary modes of description. The imaginary part of the exponent describes an oscillatory contribution, while the real part describes a relaxation process. The general case of a combination of the two represents a damped oscillation. It is illuminating to focus on the individual components separately.

An undamped oscillation $f(t) = \exp(i\omega t)$ is clearly time-translation invariant (modulo phase), since $f(t) = f(t + \Delta t)$ for $\Delta t = 2\pi n/\omega$ and $n = 1, 2, \ldots$. However, a relaxation process $f(t) = \exp(-\alpha t)$ is not. Its integration requires that specific initial conditions, the state of the system at t_{initial}, are taken into account. Hence, t_{initial} is not arbitrary and time-translation invariance is broken. This, then, provides the temporal reference point needed for the additional breakdown of time-reversal invariance.

These observations lead to a decisive criterion for reasonable candidates of neural time-keeping mechanisms. If they are intended to serve as faithful correlates of tensed time, they must include relaxation processes. Undamped oscillations alone, i.e., strictly (multi-) periodic internal clocks, do not satisfy this criterion. Only if they are coupled with a counting mechanism for n, which again requires an integrating relaxation mechanism, are they capable of exhibiting features of irreversibility required for tensed time.

The "klepsydra model" by Wackermann and Ehm (2006) is a paradigmatic example of a model which meets the criterion of relaxation without additional ancillary mechanisms. From the conceptual perspective outlined above it is, therefore, a particularly promising theoretical model of neural time-keeping. Taking two interacting klepsydrae into account, it has been developed as far as to properly match empirical results from time reproduction experiments and determine phenomenological parameters typical for the relaxation mechanisms.

Moreover, the stochastic version of the model offers the option to introduce a time operator related to the relaxation properties of stochastic (and chaotic) systems. Such a time operator naturally links the notion of irreversibility with the notion of a now with finite extension.

References

Alais, D. and Blake, R. (2005): *Binocular Rivalry*, MIT Press, Cambridge.

Atmanspacher, H., Römer, H., and Walach, H. (2002): Weak quantum theory: Complementarity and entanglement in physcis and beyond. *Foundations of Physics* **32**, 379–406.

Atmanspacher, H., Filk, T., and Römer, H. (2004): Quantum Zeno features of bistable perception. *Biological Cybernetics* **90**, 33-40.

Atmanspacher, H., Bach, M., Filk, T., Kornmeier J. and Römer, H. (2008a): Cognitive time scales in a Necker-Zeno model for bistable perception. Manuscript available at `www.igpp.de/english/tda/pdf/neckerzeno.pdf`.

Atmanspacher, H., Filk, T., and Römer, H. (2008b): Complementarity in bistable perception. *This volume.*

Basar-Eroglu, C., Strüber, D., Stadler, M., and Kruse, E. (1993): Multistable visual perception induces a slow positive EEG wave. *International Journal of Neuroscience* **73**, 139–151.

Beshkar, M. (2008): Animal consciousness. *Journal of Consciousness Studies* **15**(3), 5–33.

Chalmers, D. (2000): What is a neural correlate of consciousness ? In: Metzinger, T. (ed.), *Neural Correlates of Consciousness*. MIT Press, Cambridge, pp. 17–39.

Carter, O.L., Presti, D.E., Callistemon, C., Ungerer, Y., Liu, G.B., and Pettigrew, J.D. (2005): Meditation alters perceptual rivalry in Tibetan Buddhist monks. *Current Biology* **15**, R412–R413.

Damasio, A.R. (1999): *The Feeling of What Happens*, Harcourt Brace, New York.

Denbigh, K.G. (1981): *Three Concepts of Time*, Springer, Berlin.

Edelman, D.B., Baars, B.J. and Seth, A.K. (2005): Identifying hallmarks of consciousness in non-mammalian species. *Consciousness and Cognition* **14**, 169–187.

Edelman, G.M and Tononi, G. (2000): *A Universe of Consciousness*, Basic Books, New York.

Ee, R. van (2005): Dynamics of perceptual bi-stability for stereoscopic slant rivalry and a comparison with grating, house-face, and Necker cube rivalry. *Vision Research* **45**, 29–40.

Ee, R. van, Dam L.C.J. van, and Brouwer G.J. (2005): Voluntary control and the dynamics of perceptual bistability. *Vision Research* **45**, 41–55.

Franck, G. (2004): Mental presence and the temporal present. In: Globus, G.G., Pribram, K.H. and Vitiello, G. (eds.), *Brain and Being*. Benjamins, Amsterdam, pp. 47–68.

Franck, G. (2008): Presence and reality: An option to specify panpsychism ? *Mind and Matter* **6**, 123–140.

Gelfand, I. and Neumark, M. (1947): Unitary representations of the group of linear transformations of the straight line. *Comptes Rendus de l'Académie des Sciences de l'URSS* **55**, 567–570.

Grünbaum, A. (1963): *Philosophical Problems of Space and Time*, Knopf, New York.

Husserl, E. (1996): *The Phenomenology of Internal Time Consciousness*, Indiana University Press, Bloomington.

Jammer, M. (2006): *Concepts of Simultaneity*, Johns Hopkins, New York.

Klose, J. (2008): Process ontology from Whitehead to quantum physics. *This volume.*

Meng, M. and Tong, F. (2004): Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *Journal of Vision* **4**, 539–551.

Metzinger, T. (2003): *Being No One*, MIT Press, Cambridge.

Misra, B. and Sudarshan, E.C.G. (1977): The Zeno's paradox in quantum theory. *Journal of Mathematical Physics* **18**, 756–763.

Necker, L.A. (1832): Observations on some remarkable phenomenon which occurs in viewing a figure of a crystal or geometrical solid. *The London and Edinburgh Philosophy Magazine and Journal of Science* **3**, 329–337.

Neumann, J. von (1932): *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin.

Pashler, H.F. (1998): *The Psychology of Attention*, MIT Press, Cambridge.

Pöppel, E. (1997): A hierarchical model of temporal perception. *Trends in Cognitive Science* **1**, 56–61.

Primas, H. (2003): Time-entanglement between mind and matter. *Mind and Matter* **1**, 81–119.

Primas, H. (2008): Complementarity of mind and matter. *This volume.*

Reichenbach, H. (2000): *The Direction of Time*, Dover, New York.

Reisberg, D. and O'Shaughnessy, M. (1984): Diverting subjects' concentration slows figural reversals. *Perception* **13**, 461–468.

Ruhnau, E. (1995): Time Gestalt and the observer. In: Metzinger, T. (ed.), *Conscious Experience*. Imprint Academic, Exeter, pp. 165–184.

Seager, W. and Allen-Hermanson, S. (2001): Panpsychism. *Stanford Encyclopedia of Philosophy*, ed. by E. Zalta.

Seth, A.K., Baars, B.J. and Edelmann, D.B. (2005): Criteria for consciousness in humans and other mammals. *Consciousness and Cognition* **14**, 119–139.

Skrbina, D. (2005): *Panpsychism in the West*, MIT Press, Cambridge.

Stapp, H.P. (1999): Attention, intention, and will in quantum theory. *Journal of Consciousness Studies* **6**(8-9), 143–164.

Stapp, H.P. (2007): *The Mindful Universe*, Springer, Berlin.

Strawson, G. (2006): Realistic monism: Why physicalism entails panpsychism. *Journal of Consciousness Studies* **13**(10-11), 3–31, commentaries pp. 32–183, replies pp. 184–280.

Varela, F.J. (1999): Present-time consciousness. *Journal of Consciousness Studies* **6**(2-3), 111–140.

Vickers, D. (1972). A cyclic decision model of perceptual alternation. *Perception* **1**, 31–48.

Wackermann, J. and Ehm, W. (2006): The dual klepsydra model of internal time representation and time reproduction. *Journal of Theoretical Biology* **239**, 482–493.

Whitrow, G.J. (1980): *The Natural Philosophy of Time*, Oxford University Press, Oxford.

Synchronicity, Quantum Mechanics, and Psyche

François Martin[1] and Giuliana Galli Carminati[2]

[1] Laboratoire de Physique Théorique et Hautes Energies, Universités Paris 6 et 7, Paris, France, `martin@lpthe.jussieu.fr`

[2] Adult Psychiatry Service, Department of Psychiatry, University Hospitals of Geneva, Switzerland, `Giuliana.GalliCarminati@hcuge.ch`

1 Overview

The layout of this chapter[3] is as follows: First we recall the meaning of synchronicity effects and long-distance correlations between mental states of different individuals. Then we describe delayed-choice experiments (choice of the past) with photons which demonstrate the peculiar nature of time in quantum mechanics and could provide hints for the global nature of synchronicity effects. Further, we put forward a model of quantum entanglement between minds that could explain long-distance correlations between individual minds as well as other psychological processes, with mourning as a particular example. Such a model can explain the role of the unconscious in group insight and group consciousness. It could be useful in group therapy and in group training.

2 Synchronicity Phenomena

Synchronicity phenomena are characterized by a significant coincidence which appears between a (subjective) mental state and an event occurring in the (objective) external world. The notion was introduced by the Swiss psychoanalyst Carl Gustav Jung (1947) and further studied together with Wolfgang Pauli (Jung and Pauli, 1955; see also Atmanspacher and Primas, 1996). Synchronicity effects show no causal link between the two events that are correlated.

We can distinguish two types of synchronicity phenomena. The first one is characterized by a significant coincidence between the psyche of two individuals. An example of this type is given when two subjects at a distance buy simultaneously two identical neckties without having consulted each other

---

[3] Parts of this article are published in Carminati and Martin (2008).

beforehand. The significant coincidence appears as a correlation between the psyche of the two subjects. There are many examples of such long-range correlations between subjects: twins, relatives, members of a couple, friends, ...

Another example: I have a friend, let us call him A, in honor of whom a concert is given. The third Chopin sonata, opus 58, is performed during this concert. A has another friend, let us call him B, who has not seen A for long and who is unaware of the concert given in honor of A. Now B writes to A a long letter about the third Chopin sonata at the same time. Definitely this significant coincidence shows a correlation at a distance between the psyche of A and B, more precisely between the unconscious of A and B.

The second type of synchronicity phenomena, which is closer to what was advocated by Jung, happens when the significant coincidence occurs between a mental state and a physical state. In this case the physical state is symbolically correlated to the mental state by a common meaning. They appear not necessarily simultaneously but in a short interval of time such that the coincidence appears exceptional.

Jung thought that those phenomena appear very rarely in daily life. For him synchronistic effects emanate uniquely from an activated archetype and not from a latent one. "Such phenomena happen above all in emotional situations such as death, illness or accidents ..." (translated from Jung, 1961). However, Jung said later (Jung, 1969, §938, footnote 70):

> "I must again stress the possibility that the relation between body and soul may yet be understood as a synchronistic one. Should this conjecture ever be proved, my present view that synchronicity is a relatively rare phenomenon would have to be corrected."

We think that synchronicity is not a rare phenomenon and that it can happen in everyday life. We also think that it happens not only in emotional situations such as death, illness or accidents. It may not be a permanent phenomenon, but certainly not a rare one. As Jung said, it may necessitate some archetype to be activated.

Some people think that one could use statistical methods to prove or disprove the existence of synchronistic events. But Pauli and Jung discussed the fact that there may be a possible complementarity between statistical methods and synchronistic events (Atmanspacher and Primas, 1996), to the effect that the application of statistical methods may kill synchronistic events. Let us note that synchronistic events appear independent of volition. They are numinous events which are not controlled by the will of the person who observes them.

Synchronistic events between mind and matter seem difficult to explain in terms of correlations between conscious or unconscious minds. For Jung, synchronistic events are remnants of a holistic reality – the *unus mundus*, the "one world" of the 16th century alchemist Gerhard Dorn. This *unus mundus* could be related to Plato's world of ideas. It underlies both mind and matter, as Atmanspacher and Primas (2006) write:

> "Jung's notion of a synchronicity of pairwise arranged events in the mental and the material domains, correlated by a common meaning, is tightly related to the idea of a broken symmetry of the unus mundus. The synchronistic correlation between the events can be regarded as a retrospective indication, a remnant as it were, of the unity of the archetypal reality of the unus mundus from which they emerge."

As already stressed, in a synchronicity effect there is no causal link between correlated events localized in space and in time. Synchronicity effects are global phenomena in space and time. They cannot be explained by classical physics. However, in the case of a significant coincidence appearing between the psyche of two individuals one can see an analogy with quantum entanglement. Moreover, one can possibly see synchronistic events between the mental and the material domains as a consequence of some quantum-like entanglement between mind and matter (Primas, 2003).

Along the lines of Jung and Pauli we adopt here a dualistic view of mind and matter. Mental and material domains of reality will be considered as aspects, or manifestations, of one underlying reality in which mind and matter are unseparated (Atmanspacher, 2004).

Synchronicity phenomena, especially those involving a correlation at a distance between several individuals, lead us to postulate non-localized unconscious mental states in space and time. Mental states are not exclusively localized in the human brain. They are correlated to physical states of the brain (possibly via quantum entanglement) but they are not reducible to them.

Since we are going to study the analogy between synchronistic events and quantum entanglement, we treat mental states (conscious and unconscious) as quantum states, i.e. as vectors of a Hilbert space (Baaquie and Martin, 2005).

3 Choice of the Past: The Photon Delayed-Choice Experiment

The photon delayed-choice experiment shows the peculiar nature of time in quantum mechanics. It has been conceived by John Archibald Wheeler (1978), and has been performed in laboratories (Hellmuth *et al.*, 1987; Jacques *et al.*, 2007). The experiment is described in Figure 1. An electromagnetic wave (photon beam) is divided into two parts of equal intensity by a halv-silvered, semi-transparent mirror (mirror 1). Then two reflectors deviate each of the two beams in such a way that they intersect again at some point. Next, two detectors are set on each path of the two beams, just after the crossing point. Half of the photons are recorded in one detector (d_t) while the other half is recorded in the other detector (d_r). Therefore for each detected photon we can determine which path it went.

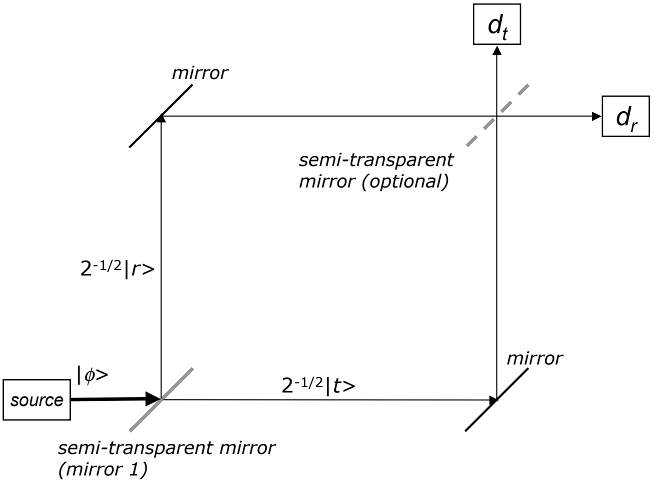


Fig. 1. The photon delayed-choice experiment; for a detailed explanation see text.

At the crossing point of the two beams we can put a second semi-transparent mirror that brings in a new phase difference between the different partial waves. The phase differences are such that all photons go into one of the detectors (d_r) and none into the other (d_t). We can choose to put, or not to put, the second semi-transparent mirror at the crossing point of the beams. Thus we can make a choice for the photon: either it follows one of the two paths when the second semi-transparent mirror is not set up, or "it follows the two paths simultaneously", so that there is an interference phenomenon, if the second semi-transparent mirror is set up at the crossing point. We can make this choice at the last moment, just before the photon reaches the crossing point, i.e. after it left the source, reached the first semi-transparent mirror and was deviated by the reflectors. We conclude that we have an effect on the past of the photon. We are able to choose the past of the photon after this past has passed by.

The delayed-choice experiment allows us to remove the indeterminacy in the past of the photon even if we act on "things" that have already happened. Wheeler stressed that "the past has no existence except as it is contained in the records, near and far, of the present". The same applies to any superposition of quantum states. Unless a measurement has been performed, or a choice has been made, such a coherent superposition of states is conserved as an indeterminacy of the past.

Quantum mechanics teaches us that there exist two levels of reality. First there is the quantum level of reality in which superposition of quantum states evolve in time in a deterministic way. For example, in the experiment described above the wave function of the photon (or the quantum electromagnetic field) evolves in a deterministic way, described by a unitary operator.

The second level of reality is what we call the level of classical reality. It is the level of what we observe with our consciousness. It is also the level that in physics is given by the result of a (single) measurement. The passage between the quantum and the classical reality is accomplished through an operation that we call "the reduction of the wave packet" (or "the collapse of the wave function"). This passage is an irreversible and non-deterministic (probabilistic) process.

In the delayed-choice experiment the wave function of the photon evolves in a deterministic way in space and time up to the point were it hits the two detectors set up on each path of the photon. The collapse of the wave function happens in the two detectors. It is probabilistic, thus non-deterministic.

When, at the crossing point of the two beams, we decide to put or not to put the second semi-transparent mirror, the past of the photon as a quantum state is fully determined. On the other hand, the state of the photon considered as a classical particle is not fully determined. The act to put or not to put the second mirror will not modify its quantum aspect before the photon reaches this mirror or the crossing point. However, it will modify the "classical" view that we have of the photon. Due to the choice that we make about the second mirror, we have an influence on the past of the photon considered as a classical system (before it reaches this mirror or the crossing point). Wheeler called this "observer-participancy". We can make choices on the classical reconstruction of the past of the photon.

If we reconstruct the classical path of the photon between the moment it has been emitted by the source and the moment it has been recorded in one of the two detectors, there must be a collapse of the wave function of the photon between those two moments. The photon delayed-choice experiment shows that this collapse happens right at the moment when the photon is recorded. Therefore there is a repercussion of the collapse of the wave function in the past. The collapse is non-local, hence global in space-time.

In agreement with this discussion, we can distinguish two types of time. First, there is the quantum time which parametrizes the evolution of any quantum state – which could be associated with the tenseless time of Primas (2003, 2008). Second, there is the classical time, the flowing time of our consciousness – which could be associated with the tensed time of Primas (2003, 2008). In quantum time, every quantum state evolves in a deterministic way; there is no indeterminacy. In classical time there is an indeterminacy of the past. It is a fundamental property of our consciousness that we are constantly reconstructing the past. Maybe we should teach ourselves to think not in classical time but in quantum time. But this is a difficult task to carry out.

In fact, if we discuss the photon delayed-choice experiment in quantum time, the photon follows the two paths in all cases, even if we do not put the second mirror. If we put it, we have the *classical illusion* that the photon has followed one of the two paths. The collapse of the wave function happens in the detectors d_t and d_r. The repercussion of the collapse in the past is also a *classical illusion.* The lesson to learn is that our interpretations of events that appear as purely classical can always be illusory.

We can imagine that something similar to the photon delayed-choice experiment happens for psychological processes. The registration of a synchronistic event by our consciousness could correspond to a collapse of the wave function which contains the potentiality of the event. Meaningful coincidences could belong to the realm of the potential, not yet actualized. They exist in the past only as potentialities, such as quantum states, such as unconscious states. They can be called phenomena only when "they are indelibly recorded by an irreversible act of amplification", i.e. by consciousness. Our acts (our choices) trigger synchronistic events such as in a delayed-choice experiment. As a consequence of such an act, a synchronistic event can appear as the collapse of a wave function that can affect a remote past (coming from a common source). It is not a local process but a global (holistic) one. This is the reason why synchronicity phenomena appear as non-causal (or a-causal).

As for the photon delayed-choice experiment, the collapse of a wave function, affecting a remote past, could be a *classical illusion* in synchronicity effects as well. There is always an illusion in the reconstruction of the past as a succession of events. This is also the case for psychological events and, especially, for those which emerge from the unconscious.

Synchronistic events appear as non-causal and paradoxical in classical time because their reconstruction by our consciousness occurs in classical time. But in quantum time they are just "there" as potentialities, as superposed states, evolving in a unitary way with all the connections among them (e.g. via quantum entanglement), waiting to become actualized or not.

4 Models of Quantum Entanglement

4.1 Is There a Collapse of the Wave Function?

Some theories of quantum measurement try to escape the problem of the collapse of the wave function; one example is the "relative state" theory by Everett (1957; see also Wheeler, 1957). Another example is the "quantum information theory" by Cerf and Adami (1977). They analyze the measurement process in quantum mechanics from the point of view of information theory as applied to quantum entanglement. In their interpretation, the measurement process is described by entropy-conserving unitary interactions. In this framework, during the measurement process, there is neither a collapse of the wave function nor are there quantum jumps.

Cerf and Adami consider a quantum object Q and a measurement device A, itself a quantum system, which they call an *ancilla.* The measurement process begins with the quantum entanglement between Q and A (first step of von Neumann's measurement process). This corresponds to the creation of an EPR (Einstein-Podolsky-Rosen) state QA that creates *super-correlations* (i.e. quantum correlations) rather than classical correlations between Q and A. Cerf and Adami (1977) state:

> "The system QA thus created is inherently quantum, and cannot reveal any classical information. To obtain the latter, we need to create classical correlations between *part* of the EPR-pair QA and *another* ancilla A', i.e., we need to observe the quantum observer."

An EPR-triplet QAA' is then created via a unitary process. This is a pure state $|QAA'\rangle$ described by the density matrix:

$$\rho_{QAA'} = |QAA'\rangle\langle QAA'|. \tag{1}$$

Then, Cerf and Adami (1977) continue:

> "Experimentally, we are *only* interested in the correlations between A and A', and *not* in the correlations between A and Q (which are unobservable anyway)... It is immediately obvious that when summing over the quantum state Q *itself*, as paradoxically as it may appear at first sight, A and A' find themselves classically correlated and in a *mixed* state."

Therefore they obtain the reduced density matrix:

$$\rho_{AA'} = \mathrm{Tr}_Q(\rho_{QAA'}). \tag{2}$$

The von Neumann entropy of the system AA' is positive, but it is compensated by a negative conditional entropy of Q (the entropy of Q when the system AA' is known). So the total entropy of the system QAA' remains zero and QAA' stays in a pure state.

It is difficult to justify how the EPR-triplet QAA' can remain in a pure state described by $|QAA'\rangle$ after the measurement. Indeed, if the measurement of the classical correlation between A and A' reveals a particular eigenvalue of an observable X, all known models of quantum measurement predict that the quantum object Q is left in the corresponding eigenstate. A choice – the choice of the measured eigenstate – has happened, corresponding to a quantum jump and a collapse of the wave function. This is not the case, however, in the quantum information theory by Cerf and Adami.

We would need to find an experimental test that could discriminate between theories of quantum measurement that implies neither the collapse of the wave function nor a quantum jump and "ordinary" theories that do suppose (or imply) a collapse of the wave function and quantum jumps. For instance, quantum decoherence models due to the interaction of a quantum object with the environment (Zurek, 1981, 1991) belong to this latter class. Let us notice that a quantum system appears as classical, i.e. exhibits classical correlations, as soon as it is quantum-entangled with another system that remains unknown. On the account of Cerf and Adami, the measured quantum object Q is unknown, whereas in quantum decoherence the environment is unknown.

In view of our interest in mental states we want to emphasize that unconscious states do not undergo a collapse of the wave function or a quantum jump. In this case, the analogue of an EPR system remains practically in the pure state in which it was before measurement. As Pitkanen (1998) says:

> "Quantum jump/state function collapse can explain the active aspect of conscious (bodily actions, etc). But can it explain the passive aspect of consciousness involving no conscious choice (sensory experience) ?"

Let us assume that quantum jumps due to passive consciousness change only the phase associated with the state function of the subsystem so that the physical state remains as such. In this case there is no collapse of the wave function of the unconscious, and there is no destruction of the quantum-entangled states of the unconscious.

Pointer-states of consciousness, i.e. states that come to be known to consciousness, are defined by the interaction of the psyche with the environment. This interaction with the environment brings to consciousness states that are compatible with the environment and thus with the classical reality that surrounds us. Pointer-states correspond to a minimum of entropy created by interaction with the environment.

4.2 Quantum Model of Mourning

Let us now try to apply the approach by Cerf and Adami to psychological processes. As a model of correlations between unconscious and conscious states we consider the case of mourning (Carminati and Carminati, 2006). Mourning is a binary situation. For example let us consider the case of Bob who has to face the death of his father. We will consider Bob's unconscious state of grief and designate it by $|BD\rangle$, a vector of a Hilbert space.

As a consequence of the interaction with the environment we will suppose that there exist two pointer-states, i.e. two stable states of which Bob can become aware. First, there would be the state $|BD1\rangle$ that would correspond to totally absent grief (Bob would not have "accepted" at all his father's death). Then there would be the state $|BD0\rangle$ for which mourning would take place (Bob would have "accepted" completely his father's death).

It seems to us that these two states can represent realistic pointer-states insofar as each of them is associated with some kind of reality. The first state is associated with the attitude that the father were still alive, while the second state is associated with the realization that the father is in fact deceased. Those two pointer-states correspond to the answers that Bob can give to the question of whether or not his father passed away. We will suppose that each of those two states is of minimal entropy as far as the interaction with the environment is concerned.

The state of Bob's unconscious related to his grief is a superposition of the two pointer-states $|BD1\rangle$ and $|BD0\rangle$. We parametrize this superposition by the angles θ and ϕ, as in the Bloch sphere representation:

$$|BD\rangle = \sin\left(\tfrac{1}{2}\theta\right)|BD0\rangle + \cos\left(\tfrac{1}{2}\theta\right)e^{i\phi}\,|BD1\rangle. \tag{3}$$

The states of consciousness corresponding to the two pointer-states, respectively, will be designated by $|BC1\rangle$ and $|BC0\rangle$. To be more precise, in fact they *are* the pointer-states.

Using the theory of Cerf and Adami, we will suppose the existence of an intermediate quantum system between Bob's unconscious state $|BD\rangle$ and Bob's conscious state $|BC\rangle$. In other words, we consider an *ancilla* which mediates the transition from the unconscious to a conscious state. We will assume that this is a kind of insight, allowing intuitions to reach our consciousness. The ancilla represents an unconscious (or preconscious) quantum system; a part of the unconscious functioning of our brain that, for Bob, we will designate by $|BI\rangle$.

In a first stage, an EPR-doublet will be formed between Bob's unconscious and Bob's insight:

$$|BD,BI\rangle = \sin\left(\tfrac{1}{2}\theta\right)|BD0\rangle|BI0\rangle + \cos\left(\tfrac{1}{2}\theta\right)e^{i\phi}\,|BD1\rangle|BI1\rangle\ , \tag{4}$$

and, in a second stage, this EPR-doublet will form an EPR-triplet with Bob's conscious state:

$$\begin{aligned}&|BD,BI,BC\rangle\ = \\ &\sin(\tfrac{1}{2}\theta)\,|BD0\rangle|BI0\rangle|BC0\rangle + \cos(\tfrac{1}{2}\theta)\,e^{i\phi}\,|BD1\rangle|BI1\rangle|BC1\rangle.\end{aligned} \tag{5}$$

This EPR-triplet is a pure state, written in the basis of pointer-states $|BC0\rangle$ and $|BC1\rangle$. It describes a quantum entanglement between the unconscious, the insight, and consciousness. The density matrix of this pure state is:

$$\rho_{BD,BI,BC} = |BD,BI,BC\rangle\langle BD,BI,BC|. \tag{6}$$

Following Cerf and Adami, we sum over the unconscious states $|BD\rangle$ to which Bob has no access and obtain the reduced density matrix:

$$\rho_{BI,BC} = \mathrm{Tr}_{BD}(\rho_{BD,BI,BC})\ , \tag{7}$$

that is to say

$$\begin{aligned}&\rho_{BI,BC}\ = \\ &\sin^2\left(\tfrac{1}{2}\theta)\right)|BI0\rangle\langle BI0||BC0\rangle\langle BC0| + \cos^2\left(\tfrac{1}{2}\theta\right)|BI1\rangle\langle BI1||BC1\rangle\langle BC1|\ ,\end{aligned} \tag{8}$$

which exhibits a classical correlation between the insight and the states of consciousness.

Since the EPR-triplet (BD,BI,BC) is in a pure state, its von Neumann entropy vanishes:

$$S(BD,BI,BC) = 0. \tag{9}$$

On the other hand, the system (BI,BC) is a statistical mixture (exhibiting classical correlations), so its von Neumann entropy is positive:

$$S(BI,BC) = -\sin^2\left(\tfrac{1}{2}\theta\right)\log\left\{\sin^2\left(\tfrac{1}{2}\theta\right)\right\} - \cos^2\left(\tfrac{1}{2}\theta\right)\log\left\{\cos^2\left(\tfrac{1}{2}\theta\right)\right\} \ . \quad (10)$$

This positive entropy is compensated by a negative *conditional* entropy, the conditional quantum entropy of Bob's unconscious state with respect to the system composed by his insight and his consciousness. These two entropies are related by

$$S(BD,BI,BC) = S(BI,BC) + S(BD|BI,BC) = 0. \quad (11)$$

The negative conditional quantum entropy is therefore equal to:

$$\begin{aligned} &S(BD|BI,BC) \ = \ -S(BI,BC) \ = \\ &\sin^2\left(\tfrac{1}{2}\theta\right)\log\left\{\sin^2\left(\tfrac{1}{2}\theta\right)\right\} + \cos^2\left(\tfrac{1}{2}\theta\right)\log\left\{\cos^2\left(\tfrac{1}{2}\theta\right)\right\}. \end{aligned} \quad (12)$$

This is the result that we obtain by applying Cerf and Adami's theory, assuming that the pointer-states of consciousness are specified by the environment.

4.3 Correlations Between Bob and Alice

When two twins buy simultaneously and at distant places two identical neckties without having consulted each other beforehand, there is a correlation at a distance between the mental states of the twins. When my friend A receives a letter of his friend B about the third Chopin sonata, B being unaware of the concert given in honor of A in which this sonata is performed, there is also a correlation at a distance between A's and B's mental states (see section 2).

There are two ways of looking at this type of correlation at a distance between mental states. The first way is to imagine that A's and B's unconscious interact via the exchange of virtual bosons, considered as quanta of a mental field (Baaquie and Martin, 2005). These virtual bosons carry the information "Chopin's third sonata" and they trigger B's unconscious. As a consequence, B writes to A a letter on Chopin's third sonata. The second way is to imagine that these long-range correlations are consequences of quantum entanglement between two mental states, quantum entanglement between A's and B's unconscious. This is the kind of correlation that we will consider here.

When A thinks about Chopin's third sonata, or when he (she) deals with a problem concerning the interpretation of this sonata, his (her) insight is in a given quantum state $|AI\rangle$. This quantum state is a pre-conscious pure state that brings to the conscious level the information "Chopin's third sonata". When B decides to write to A a letter about Chopin's third sonata, his insight is in the quantum state $|BI\rangle$, which is the same as $|AI\rangle$. From now on we will assume that A and B designate Alice and Bob, respectively.

When two twins decide, without previous agreement, to buy practically simultaneously the same necktie, their insights are also in the same quantum state, respectively. We can therefore imagine that in the situations just illustrated there is a kind of Bose-Einstein condensation that happens at the unconscious level, as well as at the level of the insight.[4] A portion of Alice's unconscious "condensates" with a portion of Bob's unconscious to form a sort of group unconscious described by a single quantum state. In a similar way, a portion of Alice's insight "condensates" with a portion of Bob's insight to form a kind of group insight also described by a single quantum state. A kind of coalescence effect happens, akin to superfluidity or superconductivity, at the unconscious and insight levels.

Nevertheless, via the continuous transition from unconscious to conscious states, the insight continuously changes its state, as consciousness itself does. Our insight is thus not always in a state of group insight. In fact, for most of the time, it is in a state of individual insight. This is the reason why the twins, or the two partners of a couple, are not continuously having the same thoughts. This is also the reason why long-range correlations do not necessarily happen exactly simultaneously. Bob did not write his letter about Chopin's third sonata at precisely the instant when Alice thought about this sonata. This does not prevent a quantum correlation between their unconscious (formation of a group unconscious) or the formation of a group insight leading to a certain form of group consciousness.

The fact that a group insight may arise without a total fusion of the consciousness of the members of the group can be compared to a superconductor where a certain number of electrons are bound into Cooper pairs. They form the superfluid (or superconducting) part of the system, while there are still "individual" electrons not bound into Cooper pairs, constituting the "normal" component of the system. In this sense, group insight is analogous to the "superfluid" component of the system, while individual insight is analogous to the "normal" component of the system.

4.4 Mourning and the Correlation Between Alice and Bob

Let us now reconsider the mentioned example, in which Bob, whose father died, sees Alice, a psychoanalyst. The state of Bob's unconscious related to the mourning process that he undergoes is given by Eq. (3). During a psychoanalysis session, Alice's unconscious $|AD\rangle$ interacts with that part of Bob's unconscious which is related to his grief, $|BD\rangle$, thus forming an EPR state described by:

$$|BD, AD\rangle \;=\; \sin\left(\tfrac{1}{2}\theta\right)|BD0\rangle|AD0\rangle + \cos\left(\tfrac{1}{2}\theta\right)e^{i\phi}|BD1\rangle|AD1\rangle. \qquad (13)$$

This is a definition of the states $|AD0\rangle$ and $|AD1\rangle$ of Alice's unconscious entangled with the unconscious mourning states of Bob. Thanks to the situation of quantum entanglement and to her insight, Alice can realize Bob's

---

[4] Fröhlich (1968) proposed a model of Bose-Einstein condensation in biological systems. This model has been adopted by Marshall (1989), where it plays a major role in establishing global brain activity. In our case, Bose-Einstein condensation is situated at the level of unconscious mental states, as opposed to the level of physical states of the brain.

mourning states. So, as far as Alice and the quantum correlation of her unconscious with Bob's unconscious are concerned, we have an EPR-quadruplet, similar to the EPR-triplet in Eq. (5):

$$\begin{aligned}|BD, AD, AI, AC\rangle \;&=\; \sin\left(\tfrac{1}{2}\theta\right)|BD0\rangle|AD0\rangle|AI0\rangle|AC0\rangle \\ &+\; \cos\left(\tfrac{1}{2}\theta\right)e^{i\phi}|BD1\rangle|AD1\rangle|AI1\rangle|AC1\rangle,\end{aligned}$$

in which $|AI\rangle$ and $|AC\rangle$ are the states of Alice's insight and Alice's consciousness, respectively. $|AI0\rangle$ and $|AC0\rangle$ are correlated with Bob's mourning state $|BD0\rangle$, and $|AI1\rangle$ and $|AC1\rangle$ are correlated with Bob's mourning state $|BD1\rangle$.

The density matrix representing the pure state $|BD, AD, AI, AC\rangle$ is

$$\rho_{BD,AD,AI,AC} \;=\; |BD, AD, AI, AC\rangle\langle BD, AD, AI, AC|. \qquad (14)$$

As we did for Bob above, we now sum over the unconscious states $|BD, AD\rangle$ to which Alice has no access, and obtain a reduced density matrix:

$$\rho_{AI,AC} = \mathrm{Tr}_{BD,AD}(\rho_{BD,AD,AI,AC}), \qquad (15)$$

that is

$$\begin{aligned}&\rho_{AI,AC} \;=\\ &\sin^2\left(\tfrac{1}{2}\theta\right))|AI0\rangle\langle AI0||AC0\rangle\langle AC0| + \cos^2\left(\tfrac{1}{2}\theta\right)|AI1\rangle\langle AI1||AC1\rangle\langle AC1|, \qquad (16)\end{aligned}$$

analogous to the reduced density matrix in Eq. (8). As for Bob, this procedure reveals a classical correlation between Alice's insight and her conscious states related to Bob's mourning.

The existence of the EPR-quadruplet $|BD, AD, AI, AC\rangle$ and of a classical correlation between Alice's insight states and her conscious states allows her to realize, at a given moment (during the analysis session), the mourning states of Bob's unconscious. Equation (16) gives the statistical weights of her thoughts "Bob has realized his mourning" or "Bob has not realized his mourning". During the analysis session, Alice can, according to the thoughts that come to her consciousness, actualize some of them via spoken language. This could help Bob to "complete" his mourning process, symbolized by an evolution of the angle θ from zero towards π, as shown in Figure 2.

Bob's mourning is represented by the unitary evolution of the entangled state $|BD, AD\rangle$ (Eq. 13) as a function of Bob's psychological time, an evolution that we can assume as adiabatic (with no variation of entropy). Thus θ will be a function of Bob's psychological time, which is linked to physical time and to Alice's psychological time. At the death of Bob's father, θ will be at zero or very close to zero (mourning has not started yet).[5] Bob is then in a state of denial or refusal. If the mourning evolves positively, θ will evolve from

---

[5] In cases in which we know that someone is going to die, the mourning may have started before his or her physical death.

zero to π, describing a consciously explicit mourning as a "healthy" development. (Note that θ does not necessarily grow monotonically as a function of psychological time.) In the case of "pathological" mourning θ may remain frozen at a value close to zero. When the value of θ is between zero and π this may be experienced as a state of depression.

Summarizing, we have applied the quantum information theory of Cerf and Adami to the process of mourning (individually or with the help of a therapist). However, our model is more general and can be applied to any psychological process involving quantum entanglement of the unconscious with insight and with consciousness. Similarly, we can describe quantum entanglement between several unconscious states.

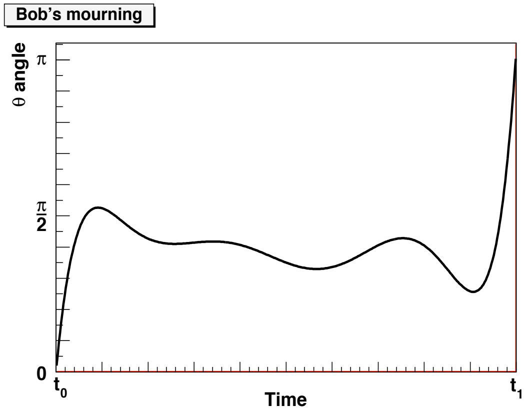


Fig. 2. Evolution of θ, the extent to which "mourning is achieved", as a function of Bob's psychological time. According to Eq. (16), $\theta = \pi$ represents the case of completely realized mourning.

4.5 Group States

The state of Alice's insight $|AI0\rangle$, which makes her realize her unconscious state $|AD0\rangle$, is the same as Bob's insight state $|BI0\rangle$, which makes him realize his unconscious state $|BD0\rangle$; $|AD0\rangle$ is quantum correlated with $|BD0\rangle$. In the same way, the state of Alice's insight $|AI1\rangle$ is the same as the state $|BI1\rangle$ of Bob's insight. We can therefore define the states $|I\rangle$ of the group insight of Bob and Alice as:

$$|I0\rangle \;=\; |BI0\rangle|AI0\rangle, \tag{17}$$

and

$$|I1\rangle \;=\; |BI1\rangle|AI1\rangle. \tag{18}$$

We can also use the unconscious states of Bob and Alice to define the states $|D\rangle$ of their group unconscious:

$$|D0\rangle \;=\; |BD0\rangle|AD0\rangle, \tag{19}$$

and

$$|D1\rangle \;=\; |BD1\rangle|AD1\rangle. \tag{20}$$

We can then rewrite Eq. (13) in group notation:

$$|D\rangle = \sin\left(\tfrac{1}{2}\theta\right)|D0\rangle + \cos\left(\tfrac{1}{2}\theta\right)e^{i\phi}|D1\rangle. \tag{21}$$

In a similar way we can define the states $|C\rangle$ of Bob's and Alice's group consciousness:

$$|C0\rangle \;=\; |BC0\rangle|AC0\rangle, \tag{22}$$

and

$$|C1\rangle \;=\; |BC1\rangle|AC1\rangle, \tag{23}$$

and write a group EPR-triplet similar to the EPR-triplet in Eq. (5):

$$|D, I, C\rangle = \sin\left(\tfrac{1}{2}\theta\right)|D0\rangle|I0\rangle|C0\rangle + \cos\left(\tfrac{1}{2}\theta\right)|D1\rangle|I1\rangle|C1\rangle. \tag{24}$$

With this notation, all results for Bob's and Alice's density matrices can be taken over to group unconscious, group insight, and group consciousness. Let us emphasize again that the thoughts reaching Bob's and Alice's consciousness are mostly individual thoughts, and only occasionally group thoughts.

The group EPR-triplet in Eq. (24), written for two persons (Bob and Alice), can be generalized for more persons, for example in group therapy or group training. In those groups the equivalent of the individual mourning would be the mourning of the group trainer (Vergopoulo, 1983). Elsewhere we proposed experiments to test the correlations between members of groups during training sessions by "absurd" questionnaires submitted to the members of the groups (Carminati and Martin, 2008).

Let us finally notice that correlations at a distance and synchronicity effects are well known phenomena that happen systematically between small groups in group therapy or group training sessions. Unlike correlations between two individuals or individual synchronicity effects, those phenomena are statistically reproducible and can be studied scientifically.

5 Conclusions

The photon delayed-choice experiment shows that a present act of a human being can cause a spatiotemporally non-local collapse of the wave function that can affect the past, even a remote past. This is inevitable when we reconstruct a quantum phenomenon in classical time – in this sense, it is a classical illusion. This makes it tempting to consider synchronicity effects as quantum effects. The acts and choices that we make do not only determine the vision that we have of the world in which we live. They can also explain synchronicity phenomena in which a (subjective) mental state coincides with an event happening in the (objective) external world. The global collapse in time could explain the apparent classical acausality of such phenomena. Synchronistic events could be explained by some sort of quantum entanglement between mind and matter.

Within a resolutely dualistic view of mind and matter (but taking also into account the correlations between mental states and the physical states of the brain) we have studied the phenomenon of quantum entanglement between mental states considered as quantum states. We emphasized the quantum entanglement between different mental states of different human beings. This could explain long-range correlations that sometimes reveal themselves between individuals such as twins, couples, friends, ...

In situations in which the interaction of the psyche with its environment is minimized, the quantum information theory of Cerf and Adami is particularly interesting. In this theory there is no collapse but a unitary evolution of the wave function (of individual or group unconscious). The quantum-entangled mental states are protected, and the unconscious is only slightly perturbed. In this framework we have modelled, via quantum entanglement, correlations between mental states of different individuals. Alternatively, one can regard these correlations as some sort of a Bose-Einstein condensation of parts of the unconscious and of insights.

This can be applied to many psychological processes, for instance to the process of mourning. We have modelled the realization (awareness) of elements of the unconscious related to mourning in cases where a subject is mourning and in cases where he (or she) receives the help of a psychoanalyst. In the latter case there is a quantum entanglement between the subjects's and the therapist's unconscious. As a consequence, we can speak of a group unconscious, of group insight (ancilla), and even of group consciousness.

We have investigated how an unconscious state related to mourning could evolve unitarily as a function of psychological time, allowing the mourning to be ultimately completed or not. This can be generalized to group dynamics taking place during group therapies and group trainings. As in the case of a pair of individuals, a group unconscious, group insight (ancilla), and even some form of group consciousness can be established.

In conclusion, if classical mechanics is unable to explain phenomena related to the psyche, especially synchronistic events, quantum theory might be a suitable framework to study those phenomena. But much remains to be worked out.

6 Acknowledgments

We thank Federico Carminati who greatly helped us in the preparation of this paper. We also thank Alain Connes who encouraged us to present this paper, and the organizers of this conference for their invitation.

References

Atmanspacher, H. (2004): Quantum approaches to consciousness. The Stanford Encyclopedia of Philosophy (Winter 2004 Edition), ed. by E.N. Zalta. Available at `plato.stanford.edu/entries/qt-consciousness/`.

Atmanspacher, H. and Primas, H. (1996): The hidden side of Wolfgang Pauli. *Journal of Consciousness Studies* **3**(2), 112–126.

Atmanspacher, H. and Primas, H. (2006): Pauli's ideas on mind and matter in the context of contemporary science. *Journal of Consciousness Studies* **13**(3), 5–50.

Baaquie, B.E. and Martin, F. (2005): Quantum psyche: Quantum field theory of the human psyche. *NeuroQuantology* **3**, 7–42.

Carminati, G.G. and Martin, F. (2008): Quantum mechanics and psyche. *Physics of Particle and Nuclei* **39**(4), in press.

Carminati, G.G. and Carminati, F. (2006): The mechanism of mourning: an anti-entropic mechanism. *NeuroQuantology* **4(2)**, 186–197.

Cerf, N.J. and Adami, C. (1977): Quantum mechanics of measurement. Manuscript available at `http://xxx.lanl.gov/abs/quant-ph/9605002v2`.

Everett, H. (1957): "Relative state" formulation of quantum mechanics. *Reviews of Modern Physics* **29**, 454–462.

Fröhlich, H. (1968): Long-range coherence and energy storage in biological systems. *International Journal of Quantum Chemistry* **2**, 641–649.

Hellmuth, T., Walter, H., Zajonc, A. and Schleich, W. (1987): Delayed-choice experiments in quantum interference. *Physical Review* **A 35**, 2532–2541.

Jacques, V., Wu, E., Grosshans, F., Treussart, F., Grangier, P., Aspect, A. and Roch, J.-F. (2007): Experimental realization of Wheeler's delayed-choice Gedanken experiment. *Science* **315**, 966–968.

Jung, C.G. (1947): Der Geist der Psychologie. In: O. Fröbe-Kapteyn (ed.), *Eranos-Jahrbuch 1946, Vol. XIV*. Rhein-Verlag, Zürich, pp. 385–490.

Jung, C.G. (1961): Ein Brief zur Frage der Synchronizität. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie* **5**, 1–9.

Jung, C.G. (1969): *The Collected Works of C.G. Jung. Volume 8. The Structure and the Dynamics of the Psyche.* 2nd edition, Princeton University Press, Princeton.

Jung, C.G. and Pauli, W. (1955): *The Interpretation of Nature and the Psyche.* Pantheon, New York. German original: *Naturerklärung und Psyche*, Rascher, Zürich, 1952.

Marshall, I.N. (1989): Consciousness and Bose-Einstein condensates. *New Ideas in Psychology* **7**, 73–83.

Pitkanen, M. (1998): Quantum approaches to consciousness. Manuscript available at `www.helsinki.fi/~matpitka/cbookI.html`.

Primas, H. (2003): Time-entanglement between mind and matter. *Mind and Matter* **1**(1), 81–119.

Primas, H. (2008): Complementarity of mind and matter. This volume.

Vergopoulo, T. (1983): La sensibilisation à la dynamique de groupe d'après W.R. Bion et S.H. Foulkes. *Médecine et Hygiène* **41**, 3149–3155.

Wheeler, J.A. (1957): Assessment of Everett's "relative state" formulation of quantum theory. *Reviews of Modern Physics* **29**, 463–465.

Wheeler, J.A. (1978): The "past" and the "delayed-choice" double-slit experiment. In: A.R. Marlow (ed.), *Mathematical Foundations of Quantum Theory.* Academic Press, New York, pp. 9–48.

Zurek, W.H. (1981): Pointer basis of quantum apparatus: into what mixture does the wave packet collapse? *Physical Reviev* **D 24**, 1516–1525.

Zurek, W.H. (1991): Decoherence and the transition from quantum to classical. *Physics Today* **44**(10), 36–44.

When Pauli Met Jung – the Path from “Three” to “Four”

Arthur I. Miller

Department of Science & Technology Studies, University College, London, UK,
a.miller@ucl.ac.uk

What happened when the brilliant but troubled scientist, Wolfgang Pauli, met the great psychologist Carl Gustav Jung ? Can Jung’s analysis of Pauli’s dreams shed further light on how Pauli made his greatest discovery: the exclusion principle ? This dramatic story is the core of my forthcoming book (Miller 2008), to appear in 2008. Here I present part of the story based on the lecture I was privileged to present at Ascona.

The roots of Pauli’s psychic life are intimately entwined with his scientific discoveries, beginning with the famous exclusion principle. In outline, Pauli’s route to the discovery of the exclusion principle was as follows: In 1913, Niels Bohr proposed a theory of the atom based on an iconic image of it as a miniscule solar system. In Bohr’s theory the electrons in an atom are restricted to occupy only certain orbits. Upon dropping from a higher to a lower orbit an electron emitted light of a certain frequency which could be detected in the laboratory as a spectral line.

Bohr’s theory achieved a number of stunning successes. But it was beset by problems, such as: Why did not every electron in an atom drop into the atom’s lowest orbit ? By 1922, Bohr had managed to “guess” the correct distribution of electrons among the allowed orbits, but offered no details.

Another problem was that when an atom was placed between the pole faces of a magnet, many of its spectral lines split into more lines called multiplets. Certain of these lines defied explanation by Bohr’s theory. Physicists referred to this situation as the “anomalous Zeeman effect”, after the physicist Pieter Zeeman who discovered it. It baffled everyone, including the twenty-two-year-old child prodigy Pauli who obsessed over it. Eventually he decided to cease work on the problem altogether, but kept up with the burgeoning literature.

In December 1924, while on the faculty at the University of Hamburg, he had two brainwaves. First of all, he wondered what relativity theory had to say about the predominant model of an atom with one free electron, called an alkali atom. To some degree this mimicked the simple hydrogen atom with which Bohr’s theory had had its greatest success. Scientists assumed that alkali atoms were made up of a closed inert core of electrons plus a single free electron which was able to participate in chemical reactions. In order to explain the anomalous Zeeman effect Bohr suggested that the inert core could be distorted in one of two ways by some force which he defined only vaguely.

Pauli found that electrons in the core moved at velocities close to that of light. According to relativity this would cause changes in their mass which should influence the spacing between multiplets. But experimental data revealed nothing like this. Pauli concluded that the inert core model was wrong.

Then he recalled a recent paper by Edward F. Stoner, a physicist at the University of Leeds in England. The state of each electron in the Bohr theory of the atom was determined by three integer numbers, called quantum numbers. This seemed reasonable insofar as each electron moved in three dimensions. Through clever manipulation of these numbers, Stoner was able to relate the number of multiplets for the single electron plus the core in an alkali atom undergoing an anomalous Zeeman effect to the number of electrons that filled up an orbit.

Pauli realized how to interpret Stoner's result in a way that went beyond the anomalous Zeeman effect. He allocated to each electron in every atom the two-valuedness of the useless core. After a great deal of angst he did this by assigning to each electron a fourth quantum number with the value 1/2, soon to be associated with the spin of the electron. The result was that orbits filled up according to the rule that no two electrons in an atom have the same *four* quantum numbers. This is Pauli's exclusion principle. It explains why Bohr's "guess" worked for the atomic structure of atoms. Even more, it explains the structure of the periodic table of the chemical elements. Soon scientists found that Pauli's exclusion principle played a role in explaining why metals are hard and how certain stars die.

But many physicists were mystified by it. Where did the exclusion principle come from ? Could it be derived from Bohr's theory ? Pauli concluded the most important paper he ever wrote with the statement that the true meaning of his discovery would not be clarified until a deeper understanding of the quantum theory was obtained.

In a nutshell, this was Pauli's scientific route to the exclusion principle. Informative as it is, it is not the whole story. I would like to understand his thought processes and examine whether his unconventional life outside physics had anything to do with his creative thinking. At the core of Pauli's discovery was the expansion of the quantum numbers needed to define the state of an electron in an atom from three to four. Pauli could not put the feelings of angst it took to make this breakthrough out of his mind. Although Pauli compartmentalized his mental activity, might it have been that the watertight compartments of his mind broke down during his creative outburst of 1924 ?

One contributing factor in this respect was his secret nocturnal life. While by day Pauli was a staid Germanic physics professor, by night he often frequented the Sankt Pauli, the notorious red light district in Hamburg. There he found the means to alleviate his personal anger and the strains put on his psyche by a life defined by his physics research which he considered a failure due to his inability to solve the anomalous Zeeman effect. He dropped into the underworld of drugs, alcohol, prostitution and pornography.

In the fall of 1927 a calamitous event occurred in Pauli's personal life – the suicide of his mother to whom he was very close. His father, Wolfgang senior, ever the womanizer, had gone one step too far. He had left his wife for another woman, a humiliation too much for her to bear. The great compartmentalizer, Wolfgang junior, never discussed his mother's death with colleagues. Instead he buried himself ever-deeper into his research. Luckily, the following year, the call to a professorship at the ETH, in Zürich, arrived offering Pauli the opportunity for a fresh start. Although in stolid Zürich there was no Sankt Pauli his travels back to Hamburg and Berlin made up for it. Socially he was managing quite well amid a congenial group of young physicists.

In the spirit of his favorite philosopher Arthur Schopenhauer, Pauli had always dismissed marriage as a bourgeois institution. So colleagues were amazed when in December 1929 he announced that he was to marry a cabaret dancer called Käthe Deppner. He had met her some years earlier during one of his many jaunts into the fetid demi-monde of Berlin. Suffice it to say that it was a mismatch. The marriage lasted less than a year.

Meanwhile, Pauli's scientific creativity never flagged. One month after the divorce, in December 1930, Pauli suggested a new particle – the neutrino – in order to preserve the conservation law of energy in beta-decay. Today new particles are suggested on an almost daily basis. In those days it was an audacious move – were not the electron, proton and light quantum enough ? It was extraordinary that at a time of such enormous personal trauma Pauli should come up with a concept of such importance. His powers of compartmentalisation were indeed astounding.

Far from taking his divorce from Käthe Deppner in the witty, sardonic way in which he presented it to others, he had gone on a binge of drinking and parties and resumed his life of bar-room brawls, smoking and womanizing. Eventually his bitter quarrels with colleagues at the ETH came to the attention of the administration, putting his position in jeopardy despite his brilliant work. He seemed to be living two separate mental lives. To add to all this, his always vivid "dreams and visions" were seeping into his waking life (Jung, 1975, §1264-§1275). By the beginning of 1932 he had plummeted to a frightening low point. Despite his difficult relationship with of him, Pauli heeded his father's advice to consult the celebrated psychoanalyst Carl Gustav Jung who, at fifty-seven years of age, was at the height of his fame.

Unlike Freud, Jung was interested in aspects of the psyche that could not be attributed to an individual's personal development but to the deeper non-personal realms common to humankind – the collective unconscious, whose contents he called "archetypes". These are not inherited ideas, rather they are latent potentialities whose origins remain forever obscure because they reside in the mysterious realm of the collective unconscious about which we will never have direct knowledge (Jung, 1969, §718). Whereas the archetype itself is not representable, its effects enable us to visualize it as an archetypal image, or symbol.

Archetypes are hard-wired into the mind and serve as organizing principles allowing us to construct knowledge from the potpourri of sensations bombarding us. They influence our thoughts, feelings and actions.

Before visiting Jung, Pauli decided to read up on his psychology. He carefully studied Jung's 1921 book *Psychological Types*, in which Jung established a vocabulary and framework for his "analytical psychology". On the basis of his clinical experience and vast knowledge of Eastern and Western religions, philosophy and literature, Jung offered a theory of the mind based on two opposing psychological types: introverts and extraverts. He fine-tuned these notions with four basic functions, which he called thinking, feeling, intuition and sensation. He separated the four functions into two groups of two: thinking and feeling, intuition and sensation.

Jung sized up Pauli immediately. He saw before him a brilliant young man whose thinking function far outweighed the feeling function, causing a severe neurosis. Pauli poured out his troubles to Jung (1968, §45; 1975, §1268). He told him about his anger, his loneliness, his drunken brawls and his problems with women.

Jung and Pauli met for a short interview in November 1932 and then eight months later began to meet as regularly as possible on Mondays at noon in Jung's home in Küsnacht, just outside Zürich. Pauli had already written up 355 dreams. By the time they concluded their sessions about a year later, he had added another 45. Jung was elated. "They contain the most marvellous series of archetypal images", he said in one of the many lectures he gave on Pauli's dreams.[1] Complying with Pauli's request, he never mentioned who the dreamer was.

Jung's method in analytic psychology was to identify a patient's dream images with those from alchemy, religion and myth taking into account the four psychological functions. A typical analytical session with Jung started with a patient telling him about a dream. From the bookshelves lining the walls of his immense library or treatment room, Jung would then take down an ancient book of alchemical images and choose an appropriate one to discuss. Thus did Jung analyze Pauli's dreams. Eventually Pauli began to draw mandalas, signalling his achieving a balanced psyche. This stage culminated in Pauli's "great vision – the vision of the world clock".

In Pauli's words:[2]

> "There is vertical and a horizontal circle, having a common centre. This is a world clock. It is supported by the black bird. The vertical circle is a blue disk with a white border divided into $4 \times 8 = 32$ partitions. A pointer rotates upon it. The horizontal circle consists of four colours. On it stand four little men with pendulums, and round about it is laid the ring that was

---

[1] Moreover, he drew them "without being told to do so", said Jung (1975, §403).

[2] See the quotation by Jung (1968, §307) and Jung (1958, §111).

> once dark and is now golden (formerly carried by the children). The ‘clock’ has three rhythms or pulses:
>
> 1. The small pulse: the pointer on the blue vertical disk advances by 1/32.
> 2. The middle pulse: one complete revolution of the pointer. At the same time the horizontal circle advances by 1/32.
> 3. The great pulse: 32 middle pulses are equal to one complete revolution of the golden ring.”

Pauli now finally understood why he had struggled with the transition from three to four when he postulated the extra quantum number that would explain the structure of the atom in Bohr’s theory with the exclusion principle. At this point Pauli’s personal struggle between the numbers three and four ceased, although at first sight this may not seem to be the case in Pauli’s “great vision”. After all, the pointer on the blue circle moves in three rhythms or pulses. However it intersects with a circle divided into four parts, divided up with four colors and inhabited by four grotesque dwarves, called Cabiri, chthonic gods, dating back to ancient Greece, whose role is to protect sailors – here they are guides into the unconscious. While the *trinity* is the pulse of the system, the thirty-two pulses result from the multiplication of 4 × 8.

Jung also points out that thirty-two is a special number in the Kabbalah, connoting wisdom (Jung, 1968, §313). It can be written as the sum of twenty-two (the number of letters in the Hebrew alphabet) and ten (the number of branches of the Sephirot tree). Jung (1958, §125) reads Pauli’s dream as a vision of a three-fold rhythm interpenetrated by a *quaternity* “so that each is contained in the other”, thereby completing the incomplete trinity.

In Jung’s experience, he tells Pauli, the conscious mind could not have forced the concept of the quaternity on the unconscious. Rather there is some psychic element present which expresses itself through the quaternity – completeness of the individual. The quaternity is an archetypal symbol.

Jung’s analysis drove home to Pauli why his struggle in going from three to four in his discovery of the exclusion principle had been so very difficult. Not only was Pauli grappling with physics; he struggled with his neurosis as well. In this instance, alchemy, as Jung folded it into his analytical psychology, provided insight into the creative moment.

In 1951 Pauli wrote to his close friend and former assistant Markus Fierz:[3]

> “My way to the Exclusion Principle had to do with the difficult transition from three to four, namely, with the necessity to ascribe to the electron a fourth degree of freedom (soon explained as ‘spin’) beyond the three translational ones – that was really the chief thing.”

---

[3] Letter 1286 from Pauli to Fierz of October 3, 1951, published in Meyenn (1996).

Jung often spoke about the case of the young intellectual scientist as a glowing example of his own lifelong belief that alchemical symbols shed light on the "development of symbols of the self" – and, he may have added, of physics, too (e.g. Jung, 1968, §323).

References

Jung, C.G. (1958): *The Collected Works of C.G. Jung. Volume 11. Psychology and Religion: West and East*. Princeton University Press, Princeton.

Jung, C.G. (1968): *The Collected Works of C.G. Jung. Volume 12. Psychology and Alchemy*. 2nd edition, Princeton University Press, Princeton.

Jung, C.G. (1969): *The Collected Works of C.G. Jung. Volume 8. The Structure and the Dynamics of the Psyche*. 2nd edition, Princeton University Press, Princeton.

Jung, C.G. (1975): *The Collected Works of C.G. Jung. Volume 18. The Symbolic Life*. Princeton University Press, Princeton.

Meyenn, K. von (1996): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil I: 1950–1952*. Springer, Berlin.

Miller, A.I. (2008): *The Strange Case of Dr. Jung and Mr. Pauli: When Physics Met Psychoanalysis*. Norton, New York, to be published.

What Is Mathematics? Pauli, Jung, and Contemporary Cognitive Science

Rafael Núñez

Department of Cognitive Science, University of California, San Diego, USA,
`nunez@cogsci.ucsd.edu`

1 Introduction

The brilliant physicist Wolfgang Pauli was seriously involved with investigations in natural philosophy. Many (often unpublished) manuscripts and an abundant correspondence with prominent scholars of his time reveal his thoughts on causality, consciousness, the relationship between physics and psyche, and the complementarity of mind and matter, among others. His writings show that he was genuinely interested in the history of human ideas, from Western scientific thought to Eastern philosophies, to alchemy and occultism. Regarding the human mind, Pauli's philosophical investigations were deeply influenced by the work of the Swiss psychiatrist Carl G. Jung, especially through Jung's notions of "archetype" and "collective unconscious". Pauli's philosophical investigations addressed core and fundamental issues, such as the nature of scientific observation and the ontology of scientific theories. For the cutting-edge physicist that Pauli was, a natural extension to these questions would have been: What is mathematics? What is the nature of such a precise conceptual apparatus that makes modern physics possible?

To our knowledge, Pauli did not address these questions directly. He was more of a user of mathematics rather than a philosopher of mathematics or a pure mathematician. From his writings, however, it is possible to infer some aspects of Pauli's views on the nature of mathematics. Keeping in mind the focus of the academic meeting that the present volume addresses – *Pauli's philosophical ideas and contemporary science* – in this chapter I intend to accomplish three things. First, I want to analyze some of Pauli's views on the nature of mathematics mainly as seen through his analysis of Kepler's scientific theories and in his rich correspondence with Jung between 1932 and 1958. We will see that, inspired by Jung's archetypes, some of Pauli's ideas appear to be idealistic (or Platonic) in the sense that they seem to take mathematical ideas to exist somewhat independently of human beings, while others seem to defend the position that mathematical ideas are man-made. Second, from the perspective of contemporary cognitive science, I want to comment on some methodological and theoretical issues in Pauli's work, focusing on the method of introspection used by most thinkers of Pauli's time, and on the relationships between the notions of Jungian archetype and of image schema as it is used in contemporary cognitive linguistics. Finally, I will briefly describe my own approach to the question of the nature of mathematics, by looking at current work in the cognitive science of mathematics and the embodied cognition of human everyday abstraction. I will defend the argument that bodily-grounded human cognitive mechanisms underlying everyday abstraction, such as image-schemas and conceptual metaphor, play a crucial role in making mathematics possible. Mathematics, then, from elementary geometry to transfinite numbers is a biologically-grounded wonderful human creation.

2 Pauli, a Mathematical Platonist ?

Mathematics is a very peculiar body of knowledge. On the one hand, it is an extraordinary conceptual system characterized by the fact that the very entities that constitute it are imaginary, idealized mental abstractions. These entities cannot be perceived directly through the senses. A Euclidean point, for instance, has only location but no extension(!), and, as such, it cannot be found anywhere in the entire universe. A Euclidean point cannot be actually *perceived* or observed through any scientific empirical method. Yet, the truth of many facts in Euclidean geometry depends on this essential imaginary property and cannot be demonstrated empirically (e.g., "only one line passes through two points"). And on the other hand, mathematics provides extremely stable inferential patterns (i.e., theorems) that, once proved, stayed proved forever. What is then the nature of such a unique body of knowledge ?

Two main schools of thought in the philosophy of mathematics stand out: Platonism and formalism. The former, following Plato's doctrine, sees mathematical entities, their truths and properties, as atemporal and immutable, transcending the existence of human beings. The latter views these entities as reducible to pure formal properties and rule-driven manipulations of meaningless symbols. Perhaps because mathematics appears to be so pristine, precise, objective, and transcendental, many mathematicians and physicists (even today) endorse a Platonic view of mathematics.

The famous logician Kurt Gödel, for instance, was a hard-core Platonist. Even the shocking results of his "incompleteness theorems" did not change his views of mathematics. Gödel had formally proved that given an axiomatic system for arithmetic, there are true arithmetical statements that cannot be proved within that system. He took this result to solidify his philosophical position, that the ultimate truth in mathematics lies beyond mundane axiomatic systems and human mathematical practices.[1]

---

[1] For a summarized and non-technical analysis of Gödel's seminal work see Hintikka (2000).

One of the presenters in the meeting on which this volume is based – the French mathematician Alain Connes – holds a very similar position of mathematical Platonism. Connes believes that, at least in what concerns basic arithmetic, there is a *réalité archaïque* (archaic reality) where certain facts about numbers are true, independently of how humans do mathematics, carry formal proofs and concoct axiom systems. As Connes puts it (Connes *et al.*, 2000, p. 14–15, my translation, emphases in the original):

> "What logic brings to us, is, above all, a means of showing the limitations of the formalized axiomatic method, that is, of logical deductions within a formal system . . . This intrinsic limitation leads to the separation of what is provable within a given logico-deductive system from what is true, and that I will call 'the archaic mathematical reality'. With this term, voluntarily imprecise but whose intuitive sense must be clear, I mean to encompass at least the vast continent of arithmetical truths . . . In other words, the formal system that one uses will never exhaust the archaic mathematical reality."

And what about Pauli ? Was he a mathematical Platonist ? Certain passages of his writings suggest that he was close to a Platonic position. In the opening section of his essay on the influence of archetypal ideas on Kepler's scientific theories he writes (Pauli, 1952; translated in Pauli, 1994, p. 220):

> "What is the nature of the bridge between the sense perceptions and the concepts ? All logical thinkers have arrived at the conclusion that pure logic is fundamentally incapable of constructing such a link. It seems most satisfactory to introduce at this point the postulate of a cosmic order independent of our choice and distinct from the world of phenomena."

Although Pauli in these passages does not directly refer to mathematics, he wonders about the nature of concepts and their relation to sense perceptions. And in order to deal with such questions he dismisses pure logic as a candidate and postulates a "cosmic order" that is independent of human beings and, most importantly, distinct from world facts. This is a kind of reality that has an ontology separate from the "world of phenomena" and transcends the existence of human beings. He then further explicates this view by referring to Plato himself, and by citing Kepler as an important figure who endorsed such a view (Pauli, 1952; translated in Pauli, 1994, p. 221):

> "The process of understanding nature as well as the happiness that man feels in understanding, that is, in the conscious realization of new knowledge, seems thus to be based on a correspondence, a 'matching' of inner images *pre-existent* in the human psyche with external objects and their behavior. This interpretation of scientific knowledge, of course, goes back to Plato, and is, as we shall see, very clearly advocated by Kepler."

Pauli was well aware that the question of the nature of the "matching" between human pre-existing inner images with objects in the external world had been at the core of philosophy of mind and scientific psychology for more than a century. In this passage he presents the issue through Kepler's eyes, explaining how he straightforwardly solved the question by invoking God and creation dogmas of Christianity (Pauli, 1952; translated in Pauli, 1994, p. 221, emphasis in the original):

> "He [Kepler] speaks in fact of ideas that are pre-existent in the mind of God and were implanted in the soul, the image of God, at the time of creation. These primary images which the soul can perceive with the aid of an innate 'instinct' are called by Kepler archetypal ('archetypalis'). Their agreement with their 'primordial images' or *archetypes* introduced into modern psychology by C.G. Jung and functioning as 'instincts of imagination' is very extensive."

This passage is quite telling. Although Pauli refers to views that Kepler had expressed more than three centuries earlier, he manages to introduce the crucial Keplerian notion of "archetypes", but this time cautiously backed-up with Jung's work in "modern psychology", which at the time of Pauli was considered to be an expression of cutting-edge empirical investigation of the mind. Now, supported by Jung's empirical psychology, and getting away from dogmatic theological arguments, Pauli moves on to close his opening section by explaining how human ideas – including scientific ones – evolve (Pauli, 1952; translated in Pauli, 1994, p. 221):

> "As ordering operators and image-formers in this world of symbolical images, the archetypes thus function as the sought-for-bridge between the sense perceptions and the ideas and are, accordingly, a necessary presupposition even for evolving a scientific theory of nature."

In this opening section, Pauli says nothing about mathematics proper, but he provides the guidelines for the essential building blocks underlying scientific discovery, namely, a process that builds on archetypes and serves as the bridge between sense perceptions and the world of ideas. The semantic content of archetypes, thus, is seen as somewhat independent of human psychological activity, that is, they reside outside of the mind. Therefore they seem to be in line with platonic thought. Such a view shows up in other places in Pauli's writings. For example, in a letter of January 7, 1948, to Fierz, Pauli writes (Meyenn, 1993, pp. 496–497):[2]

> "*The ordering and regulating factors must be placed beyond the distinction of 'physical' and 'psychic'* – as Plato's 'ideas' share the notion of a concept and of a force of nature (they create actions out of themselves). I am very much in favor of referring to the 'ordering' and 'regulating' factors in terms of 'archetypes'; but then it would be inadmissible to *define* them as contents of the *psyche*. The mentioned inner images ('dominant features of the collective unconscious' after Jung) are rather *psychic* manifestations of the archetypes which, *however*, would *also* have to put forth, create, condition anything lawlike in the behavior of the corporeal world. The laws of this world would then be the *physical manifestations of the archetypes*. . . . *Each* law of nature should then have an inner correspondence and vice versa, even though this is not always directly visible today."

---

[2] I want to thank Harald Atmanspacher for pointing me to this quote.

Here Pauli states, in strong terms, that archetypes are not contents of the psyche but, rather, the inner images are psychic manifestations of them. A particular and fairly simple area of mathematics, where the relationship between the physical and the psychical can be studied under the concept of archetypes, is that of numbers. In a letter of October 24, 1953, to Pauli, Jung states that the natural numbers are the simplest of all archetypes (translation in Meier, 2001, p. 127, emphasis in the original):

> "These [the natural numbers] seem to be the simplest and most elementary of all archetypes. That they are archetypes emerges from the psychological fact that simple whole *numbers*, given the chance, amplify themselves immediately and freely through *mythological statements*; e.g. 1 = the One, absolute, nondivisible ... and thus the unconscious, the beginning, God, etc. 2 = the division of the One, the pair, the connection, the difference (agens-patiens, masculine-feminine, etc.), counting, etc. 3 = the renaissance of the One from the Two, the son, the first masculine number, etc."

Pauli, who in is early education had been in touch with the Pythagorean view that man is able to contemplate the numerical proportions of nature thanks to the inherent sense of harmony and beauty of the soul (Gieser, 2005), seems to have accepted this view. He saw mathematics as based on the archetype of numbers, and as a genuine symbolic description of reality, to the point that it can also express mental processes – including dreams – in detail (Gieser, 2005, pp. 309–310). In his letters to Jung, Pauli often describes and analyzes his own dreams in terms of archetypes and numbers, sometimes expressed in quaternarian and trinitarian structures.[3] And in other texts, such as in his "background physics" (Meier, 2001, p. 179–196) he analyzes his dreams by invoking the well-defined imaginary unit $i = \sqrt{-1}$ as a symbol not contained in the real numbers. Pauli interprets it as having the function to unite a pair of opposites and thus produce wholeness. For Pauli, mathematical representations were indeed symbolic descriptions *par excellence* (Meier, 2001, p. 195), but the mathematical entities themselves existed outside the human mind.

But there is more. Pauli was also aware that, beyond the realm of numbers, many areas of mathematics seem to be humanly developed. In his essay on Kepler's work, for instance, he is very cautious in not blindly embarking in a fully timeless, idealistic, and absolute view of mathematics. He writes (Pauli, 1952; translated in Pauli, 1994, p. 229):

> "When *Kepler* says, however, that in the Mind of God it has been eternally true that, for example, the square of the side of a square equals half the square of its diagonal, we do not, to be sure, begrudge one of the first joyful discoverers of quantitative, mathematically formulated natural laws his elation but must, as modern men, remark in criticism that the axioms of Euclidean geometry are not the only possible ones. ... I entirely share the

---

[3] This is a rich topic whose proper treatment goes beyond the scope of this chapter.

> opinion that man has an instinctive tendency, not rooted merely in external experience, to interpret his sensory perceptions in terms of Euclidean geometry. It took a special intellectual effort to recognize the fact that the assumptions of Euclidean geometry are not the only possible ones."

Here we see that Pauli does not ascribe the same "Platonic" status to Euclidean geometry (with its Platonic solids and so on) and to other forms of geometry. He clearly states that other geometries are indeed possible, and specifically points out that they are made possible by human intellectual effort. From this perspective then, according to Pauli, not *all* of mathematics would pre-exist human beings. Some domains of mathematics would be the result of the activity of the human mind. A view along these lines can also be seen in the letter of December, 12, 1950, of Pauli to Jung (translated in Meier, 2001, p. 64) in which he mentions issues regarding the foundations of mathematics, a domain of basic research in mathematics that was very active throughout Pauli's life:

> "It should be noted that the specialized field 'Fundamentals of Mathematics' is in a state of great confusion at the moment as a result of a large-scale undertaking to deal with these questions, an endeavor that failed because it was one-sided and divorced from nature. In this field of research into the fundamentals of mathematics, the 'basis of mathematical probability calculus' marks a particular low point. ... A psychological approach would be both appropriate and very useful here."

In this passage Pauli is most likely criticizing the excessive meaningless formalisms ("divorced from nature") that drove most efforts for settling the foundations of mathematics during the 20th century, and calls for an approach that brings in the richness of the human mind. Pauli's view is certainly far from the mainstream set-theoretical approaches that were *à la mode* at that time. It was much closer to Poincaré's views that saw – unlike the analytical philosophy of Frege or Russell – a strong connection between epistemology and psychology.

So, was Pauli a mathematical Platonist ? Based on the documents we have, it appears that there is no straightforward answer to the question. Or at least, no simple answer that would apply to all of mathematics. Perhaps, Pauli had a view along the lines of Gödel or Connes, that sees most mathematical practices as human – creating axiomatic systems and formal definitions, conceiving symbols and formal proofs – but in what concerns the domain of natural numbers and simple arithmetic seeing an ultimate realm of mathematical truths transcending the human mind. Perhaps Pauli, following Jung, did see something unique in whole numbers. In a letter to Pauli of October 24, 1953 (Meier, 2001, p. 127), Jung wrote that they

> "possess that characteristic of the psychoid archetype in classical form – namely, that *they are as much inside as outside.* Thus, one can never make out whether they have been *devised* or *discovered*; as numbers they are *inside* and as quantity they are *outside.*"

3 Introspection and Archetypes: A Cognitive Science Perspective

In this section I would like to briefly comment – from the perspective of contemporary cognitive science – on two aspects of Pauli's ideas about mathematics: (1) the method of introspection that he used, and (2) the notion of archetype and its relation to image schemas.

3.1 Introspection as a Method of Investigation

In their investigations about the nature of ideas and the properties of the mind, scholars of the time of Pauli and Jung approached these issues heavily relying on the method of *introspection*. They gained insight into the functioning of the mind through the *conscious* examination of their own thoughts, perception, and intuition. Since the time of Greek philosophers, introspection has played a major role in the study of the human mind. Introspection, after all, is a readily available method of investigation, practical and instantaneous, that does not require sophisticated equipment or training. In the philosophy of mathematics, various influential mathematicians of the late 19th and early 20th centuries, such as Richard Dedekind, Georg Cantor, David Hilbert, Henri Poincaré, and Hermann Weyl, developed their philosophical work mainly using introspection as a method of inquiry. They all considered, in one way or another, human intuition as a fundamental starting point for their philosophical investigations: intuitions of small integers, intuitions of collections, intuitions of movement in space, and so on (see Dedekind, 1888; Dauben, 1979, on Cantor; Kitcher, 1976, on Hilbert; Poincaré, 1913; Weyl, 1918). They regarded these fundamental intuitions of the human mind as stable and profound enough to serve as a basis for mathematics.[4] Pauli was aware of their work, and he was especially tuned into Weyl's and Poincaré's philosophical viewpoints (Gieser, 2005).

Pauli's philosophical insights, as well as those from these mathematicians, give us many important elements regarding the personal impressions these scholars had about the nature of mathematics – from the qualitative impressions of having a mathematical insight, to the description of the structure of basic intuitions, to the organization of dreams. But beyond the philosophical and historical interest that these insights may have, they present important limitations when seen from the perspective of nowadays' scientific standards.

---

[4] However, they did not think of these intuitions and basic ideas as being "rigorous" enough. This was a major reason why, later, formalism would explicitly eliminate ideas, and go on to dominate the foundational debates. Unfortunately, at that time philosophers and mathematicians did not have the scientific and theoretical tools we have today to see that human intuitions and ideas are indeed very precise and rigorous, and that therefore the problems they were facing did not have to do with a lack of rigor of ideas and intuitions. For details see Núñez and Lakoff (1998) and Lakoff and Núñez (2000).

- First, mathematicians of the time of Pauli were professionally trained to do mathematics, not to study ideas and intuitions. And their discipline, mathematics (as such), does not study ideas or intuitions. Today, the study of ideas (concepts and intuitions) in itself is a scientific subject matter, and it is not the vague and elusive philosophical object that it was at the time of Pauli. We will come back to this point in the next section.
- Second, as pointed out above, the methodology they used was mainly *introspection* – the subjective investigation of one's own impressions, feelings, and thoughts. Now we know from substantial evidence in the scientific study of intuition and cognition, that there are fundamental aspects of mental activity that are unconscious in nature and therefore inaccessible to introspection.

Indeed, thanks to the scientific investigation of the human mind, today we know that the method of introspection not only is highly unreliable but also extremely limited in scope. In terms of time scales, introspection as such requires the integration of many cognitive functions at once – attention, perception, memory, and even language – which occur at the time scale of several hundred milliseconds, seconds and minutes. This means that introspection is unable to see anything that occurs below that time scale (e.g., in a few tens of milliseconds), thus missing essential mental processes that take place within those short time frames.

Then there is the neural dynamics underlying attention, perception, memory and so on. These neural dynamics cannot be perceived directly via introspection. We simply cannot say anything about the underlying neural dynamics involved in, say, the recognition of the face of an old friend. We may have impressions and thoughts about it, but via introspection we are completely blind to the properties of the neural dynamics that make face recognition possible. Regarding memory, study after study shows that what we remember is highly unreliable, and therefore introspection applied to memories is likely to be biased by the unreliable nature of memory (see Schacter, 1996).

And then, there is the huge amount of phenomena that co-occurs with mental activity but that is outside of conscious awareness (required for introspection to take place), such as eye saccades and speech-gesture coordination, which modern cognitive science recognizes as important indicators of human thinking in real-time (McNeill, 1992; Núñez, 2006).

In sum, philosophical inquiry based mainly on introspection – although very important – gives, at best, a very limited and often biased picture of the conceptual structure that makes mathematics possible. If we want to address the question of the nature of mathematics, introspection is not the right method to do so. This applies to Pauli's (and Jung's) philosophical work.

3.2 Archetypes and Image Schemas

Pauli often cited Jung's archetypes in his writings. The following are examples of Jung's views from the 1920s and 1930s:

- "The primordeal image, elsewhere also termed *archetype*, is always collective, i.e., it is at least common to entire peoples or epochs ... [It] is the precursor of the *idea* and its matrix." (Original 1921, translation from Jung, 1971, §747 and §750.)
- "Archetypes are typical modes of apprehension." (Original 1919, translation from Jung, 1969, §280.)
- "The archetypal motifs presumably derive from patterns of the human mind that are transmitted not only by tradition and migration but also by heredity. The latter hypothesis is indispensable, since even complicated archetypal images can be reproduced spontaneously without there being any possibility of direct tradition." (Original 1937, translation from Jung, 1958, §88.)
- "I suppose ... the inherited quality to be something like the formal possibility of producing the same ideas over and over again. I have called this the 'archetype'. Accordingly, the archetype would be a structural quality or condition peculiar to a psyche that is somehow connected with the brain." (Original 1937, translation from Jung, 1958, §165.)

The notion of archetype is deep. It could have had a much greater impact in the study of the human mind of the 20th century if it had not been so difficult to investigate it empirically. The psychology of the 1950s and 1960s, dominated by behaviorism, and that of the 1970s and 1980s, dominated by the information-processing paradigm, simply did not have room for archetypes – a notion too difficult to operationalize and to encompass within strictly individualistic rule-driven views of the human mind. Interestingly, however, certain important aspects of the notion of archetypes as described by the above citations resonate in contemporary cognitive semantics, especially in what concerns the notion of *image schemas*.

Image schemas constitute an important finding in contemporary cognitive linguistics, showing that human conceptual systems can be ultimately decomposed into primitive concepts of spatial relations. Image schemas are basic dynamic topological and orientation structures that characterize spatial inferences and link language to visual-motor experience (Johnson, 1987; Lakoff and Johnson, 1999). Image schemas, like archetypes, have a specific "structural quality" that, as Jung put it, is "somehow connected with the brain" since they appear to be realized neurally, using brain mechanisms such as topographic maps of the visual field, center-surround receptive fields, gating circuitry, and so on (Regier, 1996). Moreover, they are quite close to Jung's original idea of a "primordeal image", the "precursor" of an idea and its "matrix", with a "collective" nature.

Image schemas can be studied empirically through language (and spontaneous gestures), in particular through the linguistic manifestation of spatial relations. Every language has a system of spatial relations, though they differ radically from language to language. In English, for instance, there are prepositions like in, on, through, above, and so on. Other languages have systems that often differ radically from the English system. However, the spatial relations in a given language decompose into conceptual primitives (image schemas) that appear to be universal, that is, they are "typical modes of apprehension" that are "common to entire peoples or epochs", very much like Jung's archetypes. For example, the English word "on", in the sense used in "the book is on the desk" is a composite of three primitive image schemas:

- the *Above Schema* (the book is *above* the desk),
- the *Contact Schema* (the book is *in contact* with the desk),
- the *Support Schema* (the book is *supported* by the desk).

The *Above Schema* is orientational: It specifies an orientation in space relative to the gravitational pull one feels on one's body. The *Contact Schema* is one of a number of topological schemas: It indicates an absence of a gap. The *Support Schema* is force-dynamic in nature: It indicates the direction and nature of a force. In general, static image schemas fall into one of these categories: orientational, topological, and force-dynamic. In other languages, the primitives may combine in very different ways. Not all languages have a single concept like *on* in English. For instance, even in a language as close as German, the *on* in *on* the table is rendered as *auf*, while the *on* in *on the wall* (which does not contain the Above Schema) is translated as *an*.

A common image schema that is of great importance in mathematics is the *Container Schema* (Lakoff and Núñez, 2000), which in everyday cognition occurs as the central part of the meaning of words like *in* and *out*. The Container Schema has three parts: an Interior, a Boundary, and an Exterior. This structure forms a Gestalt, in the sense that the parts make no sense without the whole. There is no Interior without a Boundary and an Exterior, no Exterior without a Boundary and an Interior, and no Boundary without sides, in this case an Inside and an Outside. This structure is topological in the sense that the boundary can be made larger, smaller, or distorted and still remains the boundary of a Container Schema.

Image schemas have a special cognitive function: they are both *per*ceptual and *con*ceptual in nature. As such, they provide a bridge between language and reasoning on the one hand and vision on the other. Image schemas can fit visual perception, as when we see the milk as being *in* the glass. They can also be imposed on visual scenes, as when we see the bees swarming *in* the garden, where there is no physical container that the bees are in. Because terms of spatial relations in a given language name complex image schemas, image schemas are the link between language and spatial perception, forming, like Jung's archetypes, "patterns of the human mind that are transmitted not only by tradition and migration but also by heredity".

As we will see in the next section, an extremely important feature of image schemas is that their *inferential structure* is *preserved* under metaphorical mappings. This feature will turn out to be a crucial component that helps bringing mathematical ideas into being.

4 Mathematics as a Product of the Embodied Human Mind

As we saw earlier, mathematics is a peculiar body of knowledge, whose objects are idealized imaginary entities. Beyond the Euclidean point, we can see the imaginary (but precise nature) of mathematics even clearer if we look at infinity where, because of the finite nature of our bodies and brains, no direct experience can exist with the infinite itself. Yet, infinity is at the core of mathematics. It lies at the very basis of many fundamental concepts such as limits, least upper bounds, point-set topology, mathematical induction, infinite sets, points at infinity in projective geometry, to mention only a few.

If mathematics is the product of human imagination, how can we explain the nature of mathematics with its unique features such as precision, objectivity, rigor, generalizability, stability, and, of course, applicability to the real world ? How can we give a cognitive account of what mathematics is, with all the precision and complexities of its theorems, axioms, formal definitions, and proofs ? And how can we do this when the subject matter is truly abstract and apparently detached from anything concrete, as in topics as transfinite numbers, abstract algebra, and hyperset theory ?

In the realm of Platonically oriented philosophies, like Gödel's or Connes', the question of the nature of mathematics does not pose a real problem, since the existence of mathematical ideas transcends the world of human ideas. This view, of course, cannot be tested scientifically and does not provide any link to current empirical work on human ideas and conceptual systems. In such Platonic views issues and questions are a matter of *faith*, not of empirical investigation. The question of the nature of mathematics does not pose major problems to purely formalist philosophies either, because in that worldview mathematics is seen as a manipulation of meaningless symbols. The question of the origin of the meaning of mathematical ideas does not even emerge in the formalist world.

In any case, any precise explanatory proposal of the nature of mathematics should give an account of the unique collection of features that make mathematics so special: precision, objectivity, rigor, generalizability, stability, and, applicability to the real world. This is what makes the scientific study of the nature of mathematics so challenging: Mathematical entities (organized ideas and stable concepts) are abstract and imaginary, yet they are realized through the biological and social peculiarities of the human animal. The challenge then is: How can a bodily-grounded view of the mind give an account of an abstract, idealized, precise, sophisticated and powerful domain of ideas if direct bodily experience with the subject matter is not possible?

In our book *Where Mathematics Comes From*, George Lakoff and I propose some preliminary answers to such questions (Lakoff and Núñez, 2000). Building on findings in mathematical cognition and the neuroscience of numerical cognition, and using mainly methods from cognitive linguistics, a branch of cognitive science, we asked: Which cognitive mechanisms are used in structuring mathematical ideas? And more specifically, which cognitive mechanisms can characterize the inferential organization observed in mathematical ideas themselves?

We suggested that most of the idealized abstract technical entities in mathematics are created via everyday human cognitive mechanisms that extend the structure of bodily experience while preserving inferential organization. Such "natural" mechanisms are, among others, image schemas and conceptual metaphors (Lakoff and Johnson, 1980; Sweetser, 1990; Lakoff, 1993; Lakoff and Núñez, 1997; Núñez and Lakoff, 2005), conceptual blends (Fauconnier and Turner, 1998, 2002; Núñez, 2005), conceptual metonymy (Lakoff and Johnson, 1980), and fictive motion (Talmy, 1988, 2003). Using a technique we called *mathematical idea analysis* we studied in detail many mathematical concepts in several areas of mathematics, from set theory to infinitesimal calculus to transfinite arithmetic. We showed how, via everyday human embodied mechanisms such as image schemas, conceptual metaphor and conceptual blending, the inferential patterns drawn from direct bodily experience in the real world get extended in very specific and precise ways to give rise to a new emergent inferential organization in purely imaginary domains. In order to see how this works, let us now take a closer look into the study of everyday conceptual mappings and inferential organization.

4.1 Conceptual Mappings and Inferential Organization

Consider the following two everyday linguistic expressions: "The spring is *ahead* of us" and "the presidential election is now *behind* us". Taken literally, these expressions do not make any sense. "The spring" is not something that can physically be "ahead" of us in any measurable or observable way, and an "election" is not something that can be physically "behind" us. Hundreds of thousands of these expressions, whose meaning is not literal but *metaphorical*, can be observed in human everyday language. They are the product of the human imagination, they convey precise meanings, and allow speakers to make precise inferences about them.

A branch of cognitive science, cognitive linguistics (and more specifically, cognitive semantics), has studied this phenomenon in detail and has shown that the semantics of these hundreds of thousands metaphorical linguistic expressions can be modeled by a relatively small number of *conceptual metaphors* (Lakoff and Johnson, 1980; Lakoff, 1993). These conceptual metaphors, which are inference-preserving cross-domain mappings, are cognitive mechanisms that allow us to project the inferential structure from a *source domain*, which usually is grounded in some form of basic bodily experience, into another one, the *target domain*, usually more abstract. A crucial component of what is modeled is inferential organization, the network of inferences that is generated via the mappings.

The above examples use quite different lexical items (i.e., one refers to a location *ahead of us*, and the other to a location *behind us*), but they are both linguistic manifestations of a single general conceptual metaphor, namely, Time Events Are Things in Unidimensional Space.[5] As in any conceptual metaphor, the inferential structure of concepts in the target domain (time, in this case) is created via a precise mapping drawn from the source domain (unidimensional space, in this case). In what concerns time expressions, for instance, cognitive linguists have identified two main forms of this general conceptual metaphor, namely, Time Passing Is Motion of an Object (which models the inferential organization of expressions such as "Chiristmas is *coming*") and Time Passing Is Motion Over A Landscape (which models the inferential organization of expressions such as "we are *approaching* the end of the month") (Lakoff, 1993).[6] The former model has a fixed canonical observer where times are seen as entities moving with respect to the observer, while the latter has times as fixed objects where the observer moves with respect to events in time.

These two forms share some fundamental features: both map (preserving transitivity) spatial locations in front of ego onto temporal events in the future, co-locations with ego onto events in the present, and locations behind ego (also preserving transitivity) onto events in the past. Spatial construals of time are, of course, much more complex, but this is basically all what we need to know here. For the purposes of this chapter, there are two very important morals to keep in mind:

a) *Truth*, when imaginary entities are concerned, is always relative to the inferential organization of the mappings involved in the underlying conceptual metaphors. For instance, "last summer" can be conceptualized as being *behind us* as long as we operate with the general conceptual metaphor Time Events Are Things in Unidimensional Space, which determines a specific bodily orientation with respect to metaphorically conceived events in time, namely, the future as being "in front of" us, and the past as being "behind" us. Núñez and Sweetser (2006), however, have shown that the details of that mapping are not universal. Through ethnographic field work, as well as cross-linguistic gestural and lexical analysis of the Aymara language of the

---

[5] Following a convention in cognitive linguistics, capitals here serve to denote the name of the conceptual mapping as such. Particular instances of these mappings, called metaphorical expressions (e.g., "she has a great future in front of her"), are not written with capitals.

[6] For a different and more recent taxonomy based on linguistic data, as well as on gestural and psychological experimental evidence, see Núñez and Sweetser (2006) and Núñez *et al.* (2006).

Andes' highlands, they provided the first well-documented case violating the postulated universality of the metaphorical orientation future-in-front-of-ego and past-behind-ego. In Aymara, for instance, "last summer" is conceptualized as being *in front of* ego, not *behind* of ego, and "next year" is not conceptualized as being *in front of* ego, but *behind* ego. Moreover, Aymara speakers not only utter these words when referring to time, but also produce co-timed corresponding gestures, strongly suggesting that these metaphorical spatial construals of time are not merely about words, but about deeper conceptual phenomena. The moral is that there is no *ultimate truth* regarding these imaginative structures. In this case, there is no ultimate truth about where, really, is the ultimate metaphorical location of the future (or the past). Truth will depend on the details of the mappings of the underlying conceptual metaphor. As we will see, this is of paramount importance when mathematical concepts are concerned: Their ultimate truth is not hidden in the structure of the universe, but it will be relative to the underlying conceptual mappings (e.g., metaphors) used to create them.

b) It is crucial to keep in mind that the abstract conceptual systems we develop are possible *because* we are biological beings with specific morphological and anatomical features. In this sense, human abstraction is *embodied* in nature. It is because we are living creatures with a salient and unambiguous front and a back, that we can build on these properties and the related bodily experiences to bring forth stable and solid concepts such as "the future in front of us". This would be impossible if we had the body of a jellyfish or of an amoeba. Moreover, abstract conceptual systems are not "simply" socially constructed, as a matter of convention. Biological properties and specificities of human bodily-grounded experience impose very strong constraints on what concepts can be created. While social conventions usually have a huge number of degrees of freedom, many human abstract concepts do not. For example, the color pattern of the Euro bills was socially constructed via convention (and so were the design patterns they have). But virtually any color ordering would have done the job. Metaphorical construals of time, on the contrary, are *only* based on a spatial source domain. This is an *empirical* observation, not an arbitrary or speculative statement: As far as we know, there is no language or culture on earth where time is conceived in terms of thermic or chromatic source domains. And there is more: not just any spatial domain does the job. Spatial construals of time are, as far as we know, always based on unidimensional space.[7] Human abstraction is thus not merely "socially constructed". It is constructed through strong non-arbitrary biological and cognitive constraints that play an essential role in constituting what human abstraction is. Human cognition is *embodied*, shaped by species-specific non-arbitrary con-

---

[7] Although they can, of course, be more complicated, e.g. in the case of cyclic or helix-like conceptions. But even in those cases the building blocks – a segment of a circle or a helix – preserve the topological properties of the uni-dimensional segment.

straints. This property is of key importance when mathematical concepts are concerned.

We are now in a position to analyze how the inferential structure of image schemas (for example, the Container Schema) is preserved under metaphorical mappings like the ones just described to generate more abstract concepts (such as the concept of Boolean class). We shall see exactly how image schemas provide the inferential structure to the source domain of the conceptual metaphor which, via the mapping, is projected onto the target domain of the metaphor to generate sophisticated mathematical concepts, in this case, Boolean-class inferences.

4.2 Structure of Image Schemas and Metaphorical Projections

When we draw illustrations of Container Schemas, we find that they look like Venn diagrams for Boolean classes. This is by no means an accident. The reason is that classes are normally conceptualized in terms of Container Schemas. For instance, we think (and speak) of elements as being *in* or *out* of a class. Venn diagrams are visual instantiations of Container Schemas. The reason that Venn diagrams work as symbolizations of classes is that classes are usually metaphorically conceptualized as containers – that is, as bounded regions in space.

Container Schemas have a logic that appears to arise from the structure of our visual and imaging system, adapted for more general use. More specifically, Container Schemas appear to be realized neurally using such brain mechanisms as topographic maps of the visual field, center-surround receptive fields, and gating circuitry (Regier, 1996). The inferential structure of these schemas can be used both for structuring space and for more abstract reason, and is projected onto our everyday conceptual system by a particular conceptual metaphor, the Classes Are Containers metaphor. This accounts for part (by no means all!) of our reasoning about conceptual categories. Boolean logic also arises from our capacity to perceive the world in terms of Container Schemas and to form mental images using them.

So, how do we normally conceptualize the intuitive pre-mathematical notion of classes ? From the perspective of mathematical idea analysis the answer is in terms of Container Schemas. In other words, we normally conceptualize a class of entities in terms of a bounded region of space, with members of the class all *inside* the bounded region and non-members outside of the bounded region. From a cognitive perspective, intuitive classes are thus metaphorical conceptual containers, characterized cognitively by a metaphorical mapping – the Classes Are Containers metaphor. Table 1 shows the corresponding mappings. This is our natural, everyday unconscious conceptual metaphor for what a class is. It grounds our concept of a class in our concept of a bounded region in space, via the conceptual apparatus of the image schema for containment. This is the way we conceptualize classes in everyday life.

Source Domain Container Schemas		Target Domain Classes
interiors of container schemas	→	classes
objects in interiors	→	class members
being an object in an interior	→	the membership relation
an interior of one container schema within a larger one	→	a subclass in a larger class
the overlap of the interiors of two container schemas	→	the intersection of two classes
the totality of the interiors of two container schemas	→	the union of two classes
the exterior of a container schemas	→	the complement of a class

Tab. 1. The metaphor CLASSES ARE CONTAINERS

We can now analyze how conceptual image schemas (in this case, Container Schemas) are the source of four fundamental inferential laws of logic. The structural constraints on Container Schemas mentioned earlier (i.e., brain mechanisms such as topographic maps of the visual field, center-surround receptive fields, gating circuitry, etc.) give them an inferential structure, which Lakoff and I called "Laws of Container Schemas" (Lakoff and Núñez, 2000). These so-called "laws" are conceptual in nature and are reflections at the cognitive level of brain structures at the neural level (see Figure 1). The four inferential laws are Container Schema versions of classical logical laws:

- *Excluded Middle.* Every object X is either *in* Container Schema A or *outside of* Container Schema A.
- *Modus Ponens*: Given two Container Schemas A and B and an object X, if A is *in* B and X is *in* A, then X is *in* B.
- *Hypothetical Syllogism*: Given three Container Schemas A, B and C, if A is *in* B and B is *in* C, then A is *in* C.
- *Modus Tollens*: Given two Container Schemas A and B and an object Y, if A is *in* B and Y is *outside of* B, then Y is *outside of* A.

Now, recall that conceptual metaphors allow the inferential structure of the source domain to be used to structure the target domain. So, the CLASSES ARE CONTAINERS metaphor maps the inferential laws given above for embodied Container Schemas (source domain) onto conceptual classes (target domain). These include both everyday classes and Boolean classes, which are metaphorical extensions of everyday classes. The entailment of such conceptual mapping is the following:

- *Excluded Middle.* Every element X is either a *member of* class A or *not a member of* class A.
- *Modus Ponens:* Given two classes A and B and an element X, if A is a *subclass* of B and X is a *member* of A, then X is a *member* of B.
- *Hypothetical Syllogism:* Given three classes A, B, and C, if A is a *subclass of* B and B is a *subclass of* C, then A is a *subclass of* C.
- *Modus Tollens:* Given two classes A and B and an element Y, if A is a *subclass of* B and Y is *not a member* of B, then Y is *not a member of* A.

The moral is that these traditional laws of logic are in fact cognitive entities and, as such, grounded in the neural structures that characterize Container Schemas. In other words, these laws are part of our bodies. Since they do not transcend our bodies, they are not laws of any transcendent reason. The truths of these traditional laws of logic are thus not dogmatic. They are true by virtue of what they mean.

4.3 Are Hypersets Sets?

Let us close this chapter by asking the following question in modern mathematics: Are hypersets sets? If not, what are they? We will see that the answer to these questions shows that mathematics is made possible by the embodied mechanisms of human imagination, such as image schemas and conceptual metaphor. Let us begin with the question: What are sets? On the formalist

Fig. 1. The "laws" of cognitive Container Schemas. The figure shows one cognitive Container Schema, A, occurring inside another, B. By inspection, one can see that, if X is in A, then X is in B, and that, if Y is outside of B, then Y is outside of A. We conceptualize physical containers in terms of cognitive containers. Cognitive Container Schemas are used not only in perception and imagination but also in conceptualization, as when we conceptualize bees as swarming *in* the garden. Container Schemas are the cognitive structures that allow us to make sense of familiar Venn diagrams.

view of the axiomatic method, a “set” is any mathematical structure that “satisfies” the axioms of set theory as written in symbols. The traditional axioms for set theory (the Zermelo-Fraenkel axioms) are often taught as being about sets conceptualized as containers. Many writers speak of sets as “containing” their members, and most students think of them that way. Even the choice of the word “member” suggests such a reading, as do the Venn diagrams used to introduce the subject. But if you look carefully through those axioms, you will find nothing in them that characterizes a container. The terms “set” and “member of” are both taken as undefined primitives. In formal mathematics, that means that they can be anything that fits the axioms. Here are the classic Zermelo-Fraenkel axioms including the axiom of choice, commonly called the ZFC axioms.

- *The axiom of extension:* Two sets are equal if and only if they have the same members. In other words, a set is uniquely determined by its members.
- *The axiom of specification:* Given a set A and a one-place predicate $P(x)$ that is either true or false for each member of A, there exists a subset of A whose members are exactly those members of A for which $P(x)$ is true.
- *The axiom of pairing:* For any two sets, there exists a set that they are both members of.
- *The axiom of union:* For every collection of sets, there is a set whose members are exactly the members of the sets of that collection.
- *The axiom of powers:* For each set A, there is a set $P(A)$ whose members are exactly the subsets of set A.
- *The axiom of infinity:* There exists a set A such that (i) the empty set is a member of A, and (ii) if x is a member of A, then the successor of x is a member of A.
- *The axiom of choice:* Given a disjointed set S whose members are nonempty sets, there exists a set C which has as its members one and only one element from each member of S.

We can see that there is absolutely nothing in these axioms that explicitly requires sets to be containers. What these axioms do, collectively, is to *create* entities called “sets”, first from elements and then from previously created sets. The axioms do not say explicitly how sets are to be conceptualized.

The point here is that, within formal mathematics, where all mathematical concepts are mapped onto set-theoretical structures, the “sets” used in these structures are not technically conceptualized as the Container Schemas we described above. They do not have container-schema structure with an interior, boundary, and exterior at all. Indeed, within formal mathematics, there are no concepts at all, and hence sets are not conceptualized as anything in particular. They are undefined entities whose only constraints are that they must “fit” the axioms. For formal logicians and model theorists, sets are those entities that fit the axioms and are used in the modeling of other branches of mathematics.

Of course, most of us do conceptualize sets in terms of Container Schemas, and that is perfectly consistent with the axioms given above. However, when we conceptualize sets as Container Schemas, a particular entailment follows automatically: *Sets cannot be members of themselves*, since containers cannot be inside themselves. Strictly speaking, this entailment does *not* follow from the axioms, but rather from our metaphorical understanding of sets in terms of containers. The above axioms do not rule out sets that contain themselves. Indeed, an extra axiom was proposed by von Neumann to rule out this possibility:

- *The axiom of foundation:* There are no infinite descending sequences of sets under the membership relation. That is, $\cdots \in S_{i+1} \in S_i \in \cdots \in S$ is ruled out.

Since allowing sets to be members of themselves would result in such a sequence, this axiom has the indirect effect of ruling out self-membership.

Within formal mathematics, model theory has nothing to do with everyday understanding. Model theorists do not depend upon our ordinary container-based concept of a set. Indeed, certain model theorists have found that our ordinary grounding metaphor that SETS[8] ARE CONTAINERS gets in the way of modeling kinds of phenomena they want to model, especially recursive phenomena. For example, take expressions like

$$x = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}.$$

If we observe carefully, we can see that the denominator of the main fraction has in fact the value defined for x itself. In other words, the above expression is equivalent to

$$x = 1 + \frac{1}{x}.$$

Such recursive expressions are common in mathematics and computer science. The possibilities for modeling such expressions using "sets" are ruled out if the only kind of "sets" used in the modeling cannot have themselves as members. Set theorists have realized that a new non-container metaphor is needed for thinking about sets, and have explicitly constructed one (see Barwise and Moss, 1991).

The idea is to use graphs, not containers, for characterizing sets. The kinds of graphs used are accessible pointed graphs, or APGs. "Pointed" indicates an asymmetric relation between nodes in the graph, indicated visually by an arrow pointing from one node to another – or from one node back to that node itself (see Figure 2). "Accessible" indicates that there is a single node which is linked to all other nodes in the graph, and can therefore be "accessed" from any other node.

---

[8] There are technical differences between classes and sets whose analysis goes beyond the scope of this text. For a discussion see Lakoff and Núñez (2000).

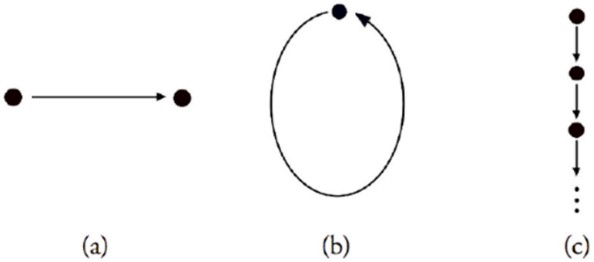


Fig. 2. Hypersets: Sets conceptualized as graphs, with the empty set as the graph with no arrows leading from it. The set containing the empty set is a graph whose root has one arrow leading to the empty set (a). Illustration (b) depicts a graph of a set that is a "member" of itself, under the Sets Are Graphs metaphor. Illustration (c) depicts an infinitely long chain of nodes in an infinite graph, which is equivalent to (b).

From the axiomatic perspective, the axiom of foundation has been replaced by another axiom that implies its negation, the "anti-foundation axiom". From the perspective of mathematical idea analysis, the creators of hypersets implicitly used a conceptual metaphor which has the mapping shown in Table 2. The effect of this metaphor is to eliminate the notion of containment from the concept of a "set". The graphs have no notion of containment built into them at all. And containment is not modeled by the graphs.

Graphs that have no loops satisfy the ZFC axioms and the axiom of foundation. They thus work just like sets conceptualized as containers. But graphs that do have loops model sets that can "have themselves as members". They do not work like sets that are conceptualized as containers, and they do not satisfy the axiom of foundation.

Source Domain Accessible Pointed Graphs		Target Domain Sets
an AGP	→	the membership structure of a set
an arrow	→	the membership relation
nodes that are tails of arrows	→	sets
decorations on nodes that are heads of arrows	→	members
AGP's with no loops	→	classical sets with the foundation axiom
AGP's with or without loops	→	hypersets with the the anti-foundation axiom

Tab. 2. The metaphor Sets Are Graphs

A “hyperset” is an APG that may or may not contain loops. Hypersets thus do not fit the axiom of foundation, but rather another axiom with the opposite intent:

- *The anti-foundation axiom:* Every APG pictures a unique set.

The fact that hypersets satisfy the Zermelo-Fraenkel axioms confirms what we said above: *The Zermelo-Fraenkel axioms for set theory – generally accepted in mathematics – do not define our ordinary concept of a set as a container.* That is, the axioms of “set theory” are not, and were never meant to be, about what we ordinarily call “sets” as conceptualized in terms of Container Schemas.

So what are sets, really? The answer to this question allows us to see the power of conceptual metaphor in mathematics. Sets, conceptualized in everyday terms as containers, do not have the right properties to model everything needed. So we can now metaphorically reconceptualize “sets” to exclude containment by using certain kinds of graphs. The only confusing thing is that this special case of graph theory is still called “set theory” for historical reasons.

Because of this misleading terminology, it is sometimes said that the theory of hypersets is “a set theory in which sets can contain themselves.” From a cognitive point of view this is completely misleading because it is not a theory of “sets” as we ordinarily understand them in terms of containment. The reason that these graph theoretical objects are called “sets” is a functional one: They play the role in modeling axioms that classical sets with the axiom of foundation used to play.

The moral is that mathematics has (at least) two internally consistent, but mutually inconsistent metaphorical conceptions of sets: one in terms of Container Schemas and one in terms of graphs. Is one of these conceptions right and the other wrong? A Platonist might want to think that there must be only one literally correct notion of a “set” transcending the human mind. But from the perspective of mathematical idea analysis these two distinct notions of a “set” define different and mutually inconsistent subject matters, conceptualized via radically different human conceptual metaphors. Mathematics is full of cases like this one.

As we mentioned at the beginning, Wolfgang Pauli in his essay on Kepler, made very clear, for the case of geometry, that “it took a special intellectual effort to recognize the fact that the assumptions of Euclidean geometry are not the only possible ones”. Perhaps we will never know what exactly Pauli meant by “intellectual effort”. Was it an effort for discovering other forms of truth in some Platonic realm? Or was it an effort in the sense of conceiving entirely new ideas thanks to cognitive mechanisms that sustain human imagination? Our work in the cognitive science of mathematics endorses the latter, which sees mathematics, from Euclidean points to transfinite numbers and hypersets, as a bodily-grounded wonderful human creation.

References

Barwise, J. and Moss, L. (1991): Hypersets. *The Mathematical Intelligencer* **13**(4), 31–41.

Connes, A., Lichnerowicz, A. and Schützenberger, M.P. (2000): *Triangle de Pensées.* Odile Jacob, Paris. English translation as *Triangle of Thought*, AMS Publications, Providence, RI, 2001.

Dauben, J.W. (1979): *Georg Cantor: His Mathematics and Philosophy of the Infinite.* Princeton University Press, Princeton.

Dedekind, R. (1888): *Was sind und was sollen die Zahlen ?* Vieweg, Braunschweig.

Fauconnier, G. and Turner, M. (1998): Conceptual integration networks. *Cognitive Science* **2**, 133–187.

Fauconnier, G. and Turner, M. (2002): *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities.* Basic Books, New York.

Gieser, S. (2005): *The Innermost Kernel. Depth Psychology and Quantum Physics. Wolfgang Pauli's Dialogue with C.G. Jung.* Springer, Berlin.

Hintikka, J. (2000): *On Gödel.* Wadsworth, Belmont, CA.

Johnson, M. (1987): *The Body in the Mind. The Bodily Basis of Meaning, Imagination, and Reason.* University of Chicago Press, Chicago.

Jung, C.G. (1958): *The Collected Works of C.G. Jung. Volume 11. Psychology and Religion: West and East.* Princeton University Press, Princeton.

Jung, C.G. (1969): *The Collected Works of C.G. Jung. Volume 8. The Structure and the Dynamics of the Psyche.* Second edition, Princeton University Press, Princeton.

Jung, C.G. (1971): *The Collected Works of C.G. Jung. Volume 6. Psychological Types.* Princeton University Press, Princeton.

Kitcher, P. (1976): Hilbert's epistemology. *Philosophy of Science* **43**, 99–115.

Lakoff, G. (1993): The contemporary theory of metaphor. In: Ortony, A. (ed.), *Metaphor and Thought.* Second edition, Cambridge University Press, New York, pp. 202–251.

Lakoff, G. and Johnson, M. (1980): *Metaphors We Live By.* University of Chicago Press, Chicago.

Lakoff, G. and Johnson, M. (1999): *Philosophy in the Flesh.* Basic Books, New York.

Lakoff, G. and Núñez, R. (1997): The metaphorical structure of mathematics: Sketching out cognitive foundations for a mind-based mathematics. In: English, L.D. (ed.), *Mathematical Reasoning: Analogies, Metaphors, and Images.* Erlbaum, Mahwah, NJ, pp. 21–89.

Lakoff, G. and Núñez, R. (2000): *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being.* Basic Books, New York.

McNeill, D. (1992): *Hand and Mind: What Gestures Reveal About Thought.* Chicago University Press, Chicago.

Meier, C.A. (ed.) (2001): *Atom and Archetype. The Pauli/Jung Letters, 1932–1958.* Princeton University Press, Princeton.

Meyenn, K. von (ed.) (1993): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band III: 1940–1949.* Springer, Berlin.

Núñez, R. (2005): Creating mathematical infinities: The beauty of transfinite cardinals. *Journal of Pragmatics* **37**, 1717–1741.

Núñez, R. (2006): Do *real* numbers really move? Language, thought, and gesture: The embodied cognitive foundations of mathematics. In: Hersh, R. (ed.), *18 Unconventional Essays on the Nature of Mathematics.* Springer, New York, pp. 160–181.

Núñez, R. and Lakoff, G. (1998): What did Weierstrass really define? The cognitive structure of natural and ε-δ continuity. *Mathematical Cognition* **4**, 85–101.

Núñez, R. and Lakoff, G. (2005): The cognitive foundations of mathematics: The role of conceptual metaphor. In: Campbell, J.I.D. (ed.), *Handbook of Mathematical Cognition.* Psychology Press, New York, pp. 109–124.

Núñez, R., Motz, B. and Teuscher, U. (2006): Time after time: The psychological reality of the ego- and time-reference-point distinction in metaphorical construals of time. *Metaphor and Symbol* **21**, 133–146.

Núñez, R. and Sweetser, E. (2006): With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science* **30**, 401–450.

Pauli, W. (1952): Der Einfluss archetypischer Vorstellungen auf die Bildung naturwissenschaftlicher Theorien bei Kepler. In: Jung, C.G., and Pauli, W. (eds.), *Naturerklärung und Psyche.* Rascher, Zürich, pp. 109–194.

Pauli, W. (1994): *Writings on Physics and Philosophy.* Edited by C.P. Enz and K. von Meyenn. Springer, Berlin.

Poincaré, H. (1913): *Dernières pensées.* Flammarion, Paris. English translation as *Mathematics and Science: Last Essays*, Dover, New York, 1963.

Regier, T. (1996): *The Human Semantic Potential.* MIT Press, Cambridge, MA.

Schacter, D. (1996): *Searching for Memory: The Brain, the Mind, and the Past.* Basic Books, New York.

Sweetser, E. (1990): *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure.* Cambridge University Press, New York.

Talmy, L. (1988): Force dynamics in language and cognition. *Cognitive Science* **12**, 49–100.

Talmy, L. (2003): *Toward a Cognitive Semantics. Volume 1: Concept Structuring Systems.* MIT Press, Cambridge, MA.

Weyl, H. (1918): *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis.* Veit, Leipzig. English translation as *The Continuum. A Critical Examination of the Foundation of Analysis*, Dover, New York, 1994.

Psychological Research on Insight Problem Solving

Michael Öllinger[1] and Günther Knoblich[2]

[1] Parmenides Center for the Study of Thinking, 80333 München, Germany,
`michael.oellinger@parmenides-foundation.org`
[2] Rutgers University, Psychology Department, Newark, NJ 07102, USA,
`knoblich@psychology.rutgers.edu`

1 Introduction

“Albert! How did you find the theory of relativity ?” Max Wertheimer, the famous Gestalt psychologist, posed this question to his friend Albert Einstein in an attempt to understand the genesis of Einstein’s groundbreaking scientific discovery (Wertheimer, 1959). Together they reconstructed the thinking processes underlying the discovery in several conversations and stumbled upon an ingenious thought experiment that Einstein had come up with. He considered it as the turning point where suddenly many open questions that had bothered his mind for a long time were easily and almost effortlessly resolved.

Imagine you are travelling in the middle of a moving train while two bolts simultaneously strike the front and the back of the train. Imagine further that there is an external observer at the embankment of the railway. Would you perceive the struck of the bolts as being simultaneous ? Would the observer ? Einstein recognized that the moving person and the observer would likely give different answers. This, in turn, led him to the crucial insight that physical measurements depend on particular frames of reference. Einstein’s postulate about the “relativity of observations” had an extreme impact on the type of explanations that were conceivable in physics (Gruber, 1995; Knoblich and Öllinger, 2006).

Although many anecdotes describe flashes of insight as coming out of the blue, this was not the case for Einstein (and probably also not for other famous scientists who made important discoveries). Einstein had, of course, a profound knowledge in classical physics and mathematics. Nevertheless, he pondered for months and even years on the problems that led to the discovery of the theory of relativity. However, his expertise, shared with many other physicists at the time, was insufficient to find the right answers. The problem was not a lack of expertise or intellectual power. The problem was that the known theories and findings had to be seen, combined, structured, or integrated in a completely new way. Einstein's thought experiment allowed him to achieve this restructuring.

Of course, not everybody is a genius. Nevertheless, research in psychology has shown that insight is a general phenomenon that can also be observed in the average person. We get stuck with a problem we are actually competent to solve, but it seems unsolvable even if we try very hard, until at some point the solution appears out of the blue. Psychological research calls the processes that lead to such insights restructuring processes. In this contribution we will provide an overview of the cognitive and neural mechanisms enabling the restructuring of problems and the resulting insights. We start with potential definitions of the term "insight" and point out the problems with such definitions. We then sketch the Gestalt psychologists' view of productive thinking that initiated psychological research on insight at the beginning of the 20th century. Next we discuss cognitive psychologists' attempts to understand the "mysterious insight phenomenon" (Bowden *et al.*, 2005) using computational models of thinking. Finally, we give an overview of current perspectives discussed in cognitive science and cognitive neuroscience.

2 Definitions of Insight

So far there is no single definition of insight in psychological research that all researchers accept (Metcalfe and Wiebe, 1987; Weisberg, 1995). However, one can identify three different dimensions that different definitions of insight focus on: a *phenomenological dimension*, a *task dimension*, and a *process dimension*.

On a *phenomenological dimension* insight can be described as a sudden, unexpected, unintended, and surprising moment where a solution pops into someone's mind. The accompanying experience is often called "aha"-experience (Bowden and Jung-Beeman, 2003; Bowden *et al.*, 2005) and is in stark contrast to other types of problem solving where problems are solved stepwise and systematically through an exhausting and laborious process. The following description from Wegner (2002, pp. 81–82) illustrates the involuntary nature of insight:

> "The happiest inconsistency between intention and action occurs when a great idea pops into mind. The 'action' in this case is the occurrence of the idea, and our tendency to say 'Eureka!' or 'Aha!' is our usual acknowledgement that this particular insight was not something we were planning in advance. Although most of us are quite willing to take credit for our good ideas, it is still true that we do not experience them as voluntary."

Wegner's description gets at the core of the paradoxical character of insight problem solving. After several conscious, laborious, and voluntary solution attempts have repeatedly failed, an unintended and unexpected idea leads to the solution of a difficult problem.

Another approach to define insight is to identify particular tasks that provoke sudden solution ideas and to contrast them with another class of problems that are more likely to provoke stepwise solutions. The focus here is on the *task dimension.* Accordingly, researchers have tried to come up with a taxonomy of insight problems, and a variety of studies tried to identify the features that characterize insight problems and distinguish them from non-insight problems. So far there is no agreement about the criteria that clearly differentiate insight problems from non-insight problems (Weisberg and Alba, 1982; Metcalfe, 1986; Metcalfe and Wiebe, 1987; Weisberg, 1992; Weisberg, 1995; Bowden and Jung-Beeman, 2003; Chronicle *et al.*, 2004; Bowden *et al.*, 2005). One criterion we find useful is the ratio between problem difficulty and the size of the problem space (all logically possible problem states). Regular problems are easy when the problem space is small and difficult when the problem space is large. In contrast, insight problems are often very difficult, although the problem space is (very) small (Knoblich *et al.*, 1999; Öllinger *et al.*, 2006). However, from a logical point of view, task-based definitions of insight are problematic, because there is always the danger that the definition becomes circular: Insight problems are problems that require insight, and insight occurs when insight problems are solved (Dominowski and Dallob, 1995).

Therefore, now most researchers use a definition of insight that is linked to particular cognitive models of insight and focus on a *process dimension.* The core assumption here is that solving insight problems involves specific processes that are not involved in stepwise solutions of problems. One guiding assumption that has driven insight research for the past 20 years is that insight involves a number of processes that change the initial problem representation (Ohlsson, 1992; Dominowski and Dallob, 1995; Knoblich *et al.*, 1999; Knoblich *et al.*, 2001; Grant and Spivey, 2003; Jones, 2003; Kershaw and Ohlsson, 2004; Knoblich *et al.*, 2005; Knoblich and Öllinger, 2006; Öllinger *et al.*, 2006). In particular, it is assumed that problem solvers initially establish inadequate problem representations that make the solution of insight problems impossible. All solution attempts repeatedly fail and problem solvers hit an impasse. To overcome such impasses the solvers' problem representation needs to change, and this representational change can be brought about through a number of different processes. The changed problem representation enhances the space of possibilities to solve the problem.

Before we describe in more detail which specific processes for restructuring are discussed in modern research on insight we should address the roots of psychological research on insight problem solving in Gestalt psychology. Not only did the Gestalt psychologists coin the term restructuring that is still central in insight research. They were also the first to systematically study insight in the psychological laboratory.

3 Insight in Gestalt Psychology

Three facts are important for the understanding of the Gestalt psychology view on insight. First, visual perception was of central concern to the Gestalt psychologists and therefore they held the view that thinking is a lot like perception (Wertheimer, 1912; Wertheimer, 1923). Their research on visual perception resulted in the famous Gestalt laws, e.g., laws of proximity, similarity, and closure (Metzger, 1986; Ash, 1998). In the Gestalt view a Gestalt is "something else" (Koffka, 1935) than the mere representation of the given physical facts. This led to the famous *credo* that the whole is more than the sum of its parts. Despite Gestalt psychology's focus on perception it also had a lasting influence on other domains of psychology like social psychology or the psychology of thinking (Wertheimer, 1959).

Second, in the Gestalt psychologists' view thinking is by definition a kind of problem solving, characterized as goal-directed behavior, that clears out existing barriers preventing the solution to problems. The underlying metaphor is that a problem is considered as a disturbed Gestalt that "asks for" being transformed into a good Gestalt ("gute Gestalt"), *the solution.* The idea is that a disturbed Gestalt exerts a kind of driving force that pulls towards the good Gestalt. The mechanism that releases and bundles this force is called *restructuring.*

Third, the Gestalt psychologists thought of themselves as a counter movement to behaviorism. They distinguished between productive thinking (good thinking) and re-productive thinking (as a bad, blind, and mindless trial-and-error strategy) and claimed that even chimpanzees would be able to solve new problems with insight (Köhler, 1921).

As described above, restructuring was the Gestalt psychologists' core concept to address thinking. Figure 1a shows an example for a problem that, according to Wertheimer (1959), requires restructuring and allows us to explain

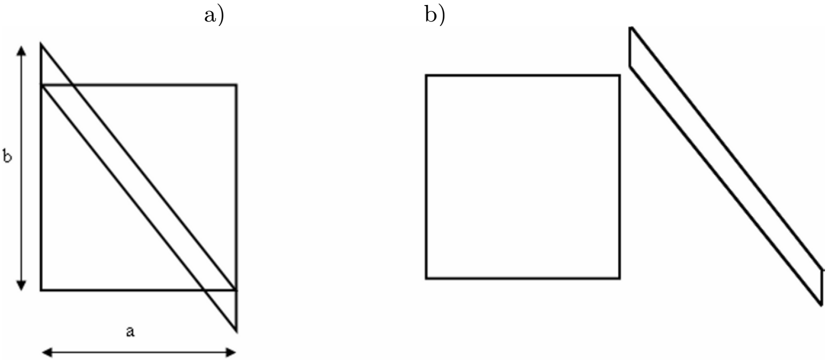


Fig. 1. The square-parallelogram problem: a) the given problem situation; b) the re-productive approach.

the difference between productive and re-productive thinking. The diagram depicts two geometrical figures, a square and a parallelogram superimposed on the square. The task is as follows: Given the lengths a and b of the two sections, what is the sum of the areas of the two figures ? If you try to find the solution by yourself you may be rewarded with the feeling of a sudden insight (but it may take some time to find it).

Most people who try to solve this problem come up with more and more complex mathematical equations by systematically varying the two given dimensions a and b. This is what their prior knowledge about the calculation of the areas of squares and parallelograms suggests. The Gestalt psychologists characterize applying one's prior knowledge in this way as re-productive thinking (Figure 1b).

However, there is a more elegant and parsimonious solution that requires productive thinking. It is illustrated at the end of this contribution in Section 10. Here the problem situation is seen from a completely new perspective. The process of restructuring leads to a rearrangement of the problem constituents that results in a much better Gestalt than the original problem (see Figure 5). The lines of the two figures are perceptually re-combined so that two rectangular triangles result. These, in turn, can be seen as a rectangle and the resulting area $a \times b$ can be easily read off. In this problem the disturbed (bad) Gestalt is literally transformed into a good Gestalt through restructuring.

Driven by their opposition against behaviorism the Gestalt psychologists believed that prior knowledge impaired productive thinking rather than supporting it. Therefore, they tried to find ways to foster productive thinking, to find out why productive thinking is often very difficult, and why people tend to "blindly" apply their prior knowledge. Karl Duncker, a disciple of Max Wertheimer and Wolfgang Köhler, investigated *functional fixedness* as a critical component that is in the way of productive thinking. In his famous candle experiment Duncker (1945) asked participants to create a ledge on the wall to rest a candle on. The given material was a candle, a matchbox, and tacks. He found that problem solvers were fixated on the "container" function of the matchbox. As a consequence they had difficulties to perceive other potential functions of the box, e.g., "support for a candle". The correct solution to the problem is to light the candle, to fix the matchbox to the wall using the tacks, to put wax on the matchbox, and to fix the candle on the box – voila!

Duncker performed further variations of this experiment. He found that presenting an empty matchbox increased the solution rate, because now the container function of the matchbox was less emphasized. Duncker made the general claim that realizing the functional-value ("Funktionalwert") of an object is the initial event that triggers successful restructuring (Duncker, 1935). For example, reaching for something that is out of one's reach, e.g., a ball under the bed requires finding a "tool" that reduces the distance to the desirable object. The functional-value of the tool is "reducing distance". Assume that an umbrella is on top of the bed. According to Duncker, two things need to happen in order for a person to use the umbrella to get the ball: First, it has to be recognized that the umbrella is a long object satisfying the required functional-value, and second, the traditional function of the umbrella "shielding against rain" has to be overcome. Further experiments confirmed Duncker's assumption (e.g., Maier, 1931) and the concept of functional fixedness is still used in cognitive psychology today.

Luchins (1942) extended Duncker's finding by demonstrating that the repeated application of the same solution procedure can result in a mental set that keeps problem solvers from finding better solutions to routine problems. Luchins defined a mental set as a state of mind that is blind for alternative and possibly easier solutions. Luchins examined mental set effects using the now famous water jug problems (Luchins, 1942; Luchins and Luchins, 1959; Lovett and Anderson, 1996). For example, given three jugs A, B, and C, with volumes of 21, 127, and 3 units, respectively, the goal might be to fill an amount of 100 units into one of the jugs. The solution to this task is to pour water into B (127), then use the water in B to fill C twice, leaving 121 units in B. The final step is to fill A using the water in B, to leave 100 units in B.

In Luchins' famous experiments participants solved a set of about two to five problems that could all be solved with the same solution procedure, B – 2C – A. Then participants were presented with a test problem that could either be solved with this solution procedure or with a simpler procedure. For example, given the volumes 23, 49, and 3 in jugs A, B, and C, with the goal of attaining 20 units, the procedure B – 2C – A can be used, but a much simpler alternative is A – C (fill A, pour once into C, and 20 units are left in A). Luchins' experiments demonstrated that participants who had used the same solution procedure on multiple problems continued to use the more complicated solution. A control group that only solved the test problems almost always applied the easier procedure. Luchins concluded that the repeated application of the same procedure makes people blind to a better approach. Of course, this was an attack against the behaviorist conviction that practice makes perfect.

To summarize, the Gestalt psychologists distinguished re-productive and productive thinking, and they thought that productive thinking was the key to make humans smart. The process of restructuring was assumed to be the core of productive thinking. This process allows problem solvers to overcome hindrances to productive thinking such as functional fixation and mental set. We will see in the next sections that the Gestalt psychologists' ideas are still very important in current research addressing insight in problem solving.

4 Cognitive Theories of Insight Problem Solving

Since the early 1960s the computer has become the dominating metaphor in research on human problem solving. After Newell and Simon (1972) published a highly influential book that conceptualized problem solving as a search in a problem space, many researchers were fairly optimistic that it would be simply a matter of time until human thinking was entirely understood and could be implemented in computational models. However, although a lot of progress has been made, today researchers are less optimistic. One reason is that computational models work best for well-defined toy problems. The mysterious nature of insight (Bowden *et al.*, 2005) remains still poorly understood. In 1986 Michael Wertheimer (the son of Max Wertheimer) raised serious doubts about whether cognitive psychology could contribute to a deeper understanding of insight problem solving (Wertheimer, 1985, p. 31):

> “ ... does modern cognitive psychology do justice to the Gestalt problem of insight ? ... from the perspective of Max Wertheimer’s book *Productive Thinking*, the answer is an unequivocal no ... It is not that the modern information-processing approaches are wrong as such; they simply do not speak to the issue of insight. They have bypassed it completely. So the basic Gestalt problem remains as unsolved and as crucial – as it was before cognitive psychology ... came on the scene.”

Although some of these doubts persist, cognitive psychologists have tried to come up with information processing models to explain insight and restructuring. Before we address some of these models we will provide an overview of the most important assumptions of problem space theory (Newell and Simon, 1972) and discuss why this theory cannot explain insight.

5 Problem Space Theory

According to problem space theory (PST), problem solving is defined as a search in a problem space. The problem space is a space of logical possibilities that is defined by an initial state (the problem), a goal state (the solution) and the operators that can be applied to a problem. The problem space encompasses all potential states that can be generated by applying the available operators to the problem, usually resulting in an exorbitant number of intermediate states that separate the initial state and the goal state. The size of the problem space depends on the given problem elements and the number of available operators. It is assumed that the problem difficulty covaries with the size of the problem space – the larger the problem space the more difficult the problem.

A classical problem that can be elegantly described and formalized by PST is the famous Tower-of-Hanoi problem (Fig. 2). The problem solver is asked to move the two disks on peg A to peg C by obeying the following rules: a) only one disk may be moved at one time and b) a larger disk may never be placed onto a smaller one. Figure 2 shows the complete problem space for the two-disk problem. If the number of disks is increased the size of the problem space explodes.

Further assumptions of PST concern the search process. Newell and Simon argued that problem solvers do not search by trial and error, randomly trying

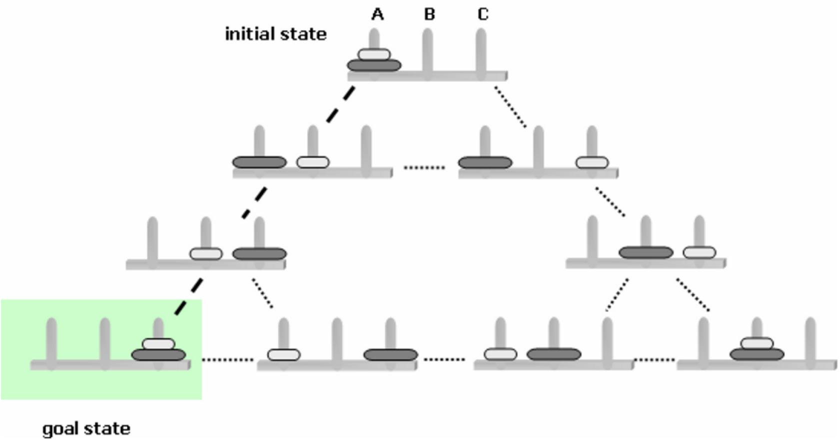


Fig. 2. The easiest version of the Tower-of-Hanoi problem with two disks. The problem space increases exponentially with the number of disks.

moves that come to mind. Instead, problem solvers apply heuristics that constrain the number of possible solution paths in a problem space. Heuristics make problem solving more efficient and more goal directed, although there is no guarantee that this will lead to a solution (Lindsay and Norman, 1981). However, heuristics can be very powerful to help to avoid searching parts of the problem space that are unlikely to contain the solution, and they help to avoid visiting the same state repeatedly. We will discuss two important heuristics, *hill climbing* and *means ends analysis*, in a bit more detail, because some accounts of insight problem solving refer to them.

The rule underlying the hill climbing heuristic is quite straightforward: Always select the move that transforms the current state into one that is as similar as possible to the goal state. This presupposes some sort of distance measure that assesses the similarity between the current state and the goal state. In the Tower-of-Hanoi problem, similarity can be defined as the number of disks that are already on peg C. This example also illustrates the problem of hill climbing – the existence of local maxima. After putting the small disk on peg C (the first move right below the initial state in Fig. 2), the current state is more similar to the goal state than the initial state, but in this constellation the solution becomes impossible, because now the larger disk can not be put on peg C. However, in many cases hill climbing is a fairly effective and parsimonious heuristic (Greeno, 1974; Thomas, 1974).

A more important heuristic is the means ends analysis (MEA). The most important characteristic of MEA is the introduction of sub-goals. MEA comprises three successive steps. In the first step the distance between initial state and goal state is determined. In the second step sub-goals are generated, and in the final step the first sub-goal that can be attained by an available operator is executed. In the Tower-of-Hanoi problem the large disk must be put onto peg C – a sub-goal is generated. To do this, it is first necessary to remove the smaller disk, that is a sub-sub-goal (moving the smaller disk from peg A to peg B) must be completed, and so on. The problem with MEA is that the number of sub-goals can become quite large which limits its usefulness if one considers the narrow capacity limitations of human working memory.

As already mentioned before, PST is hard to apply to insight problem solving because one core assumption is that problem solving proceeds in a stepwise fashion. Another problem is the implicit assumption that problem solvers generate a representation of the full problem space. Looking at human problem solving, it seems necessary to distinguish between a "subjective" problem space and an "objective" problem space. The latter unfolds all possible states in well-defined problems (like in the Tower of Hanoi). It can only be defined if all operators are known and all states can be computed. The subjective problem space of an individual problem solver can be inadequate, e.g., the wrong elements of a problem are considered. Furthermore, problem solvers may apply the wrong heuristics. Cognitive accounts of insight problem solving have explored both possibilities and they have suggested corresponding additions and modifications to PST to account for insight and restructuring. In the following, we will discuss these different accounts.

6 Heuristics and Insight

Kaplan and Simon (1990) assume that insight problems are extremely difficult, because initially they are "over-represented". In other words, they assume that many irrelevant or even misleading features and properties are incorporated into the initial problem representation whereas crucial problem aspects are omitted. In the latter case problem solvers need to change their representation of the problem space (Kaplan and Simon, 1990, p. 377):

> "Within a given problem space, the trick lies in searching for the right operator to apply next. But if no operators seem to yield progress, one must search for a new problem space to explore."

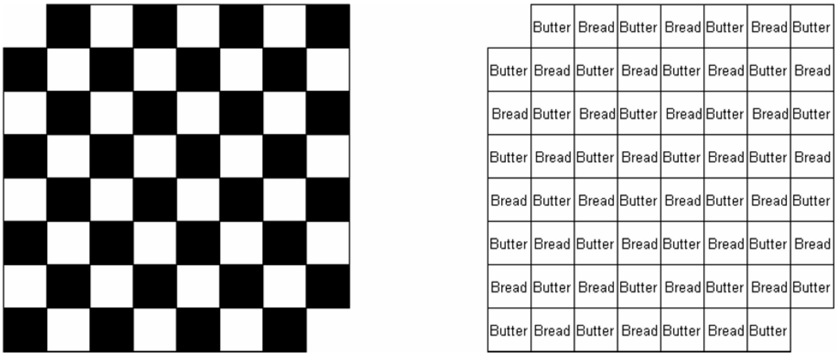
Kaplan and Simon suggest that looking for a new problem representation is a conscious process. When the problem is very difficult solvers may generate a number of different problem representations. In this case, successful problem solvers apply heuristics that enable them to detect which aspects remain invariant across different problem representations (Kaplan and Simon, 1990, p. 404):

> " . . . noticing invariants is a widely applicable rule of thumb for searching in ill-defined domains, [but] there can be no guarantee that those noticed will be the critical ones for the particular problem. Nevertheless, the constraints offered by the notice-invariant heuristic are a vast improvement over blind trial and error search."

They investigated these assumptions in a study of the mutilated checkerboard problem (Wickelgren, 1974, Figure 3). The task consists of an 8 × 8 checkerboard with two diagonally opposite corners removed. The task is to find out whether it is possible to cover the remaining 62 squares with 31 dominos, or to prove that this is impossible. A domino can cover two fields horizontally, or vertically, but not diagonally (Fig. 3). The solution is that it is impossible to cover the mutilated checkerboard with 31 dominos, because the removed corners have the same parity (same color). However, a domino can only cover two adjacent squares (black and white), and adjacent squares always have different colors.

This problem is extremely hard, even for very smart students. Only few of them were able to solve it and some of them took several days. In order to demonstrate that it is crucial to represent the parity of the two removed squares, Kaplan and Simon introduced solution hints. They found, for example, that a bread and butter version was easier to solve (Fig. 3) because this made it easier to detect that the removed fields were of the same parity ("bread").

A further theoretical approach that also emphasizes the important role of heuristics for insight problem solving is based on criteria for satisfactory progress (MacGregor *et al.*, 2001; Ormerod *et al.*, 2002; Chronicle *et al.*, 2004). The criterion for satisfactory progress theory (CSPT) postulates that successful problem solving requires two basic principles: First, problem solvers seek to maximize the consequences of each move such that the move results in a state that is as close as possible to the desired goal. This is basically the hill-climbing heuristic. Second, problem solvers constantly monitor their progress and only select moves that meet a criterion of progress – when a selected move fails to meet the criterion there is an impulse to seek alternative solutions (cf. Ormerod *et al.*, 2002, p. 792).



	Butter	Bread	Butter	Bread	Butter	Bread	Butter
Butter	Bread	Butter	Bread	Butter	Bread	Butter	Bread
Bread	Butter	Bread	Butter	Bread	Butter	Bread	Butter
Butter	Bread	Butter	Bread	Butter	Bread	Butter	Bread
Bread	Butter	Bread	Butter	Bread	Butter	Bread	Butter
Butter	Bread	Butter	Bread	Butter	Bread	Butter	Bread
Bread	Butter	Bread	Butter	Bread	Butter	Bread	Butter
Butter	Bread	Butter	Bread	Butter	Bread	Butter	

Fig. 3. The mutilated checkerboard (Wickelgreen, 1974). On the left the original problem, on the right the bread-and-butter version.

According to Ormerod and colleagues, maximization and progress monitoring together may trigger insight because they prompt the discovery and retention of so-called promising states that meet the progress-monitoring criterion. In their model previously unexpected solution paths may arise when problem solvers backtrack and start with a new "promising state". The probability of meeting an impasse (getting stuck in a solution path that does not work) also depends on a person's look-ahead. It is defined as the number of potential moves a person can consider. Of course, this varies across different individuals. People with a high look-ahead will realize more quickly that alternative ways of looking at the problem are needed because they will more quickly run out of moves that meet the progress-monitoring criterion and therefore start more quickly to look for alternative moves.

The experiments by MacGregor and colleagues on the nine-dot problem (Scheerer, 1963) illustrate the core assumptions of CSPT. The task is to connect the nine dots, arranged in a 3 × 3 matrix, with four straight lines without lifting the pen. The solution requires drawing lines beyond the boundary of the matrix (Fig. 4a). For a long time this was believed to be the main source of difficulty in this problem. However, MacGregor and colleagues think that the high difficulty is mainly due to the fact that problem solvers apply inappropriate heuristics.

According to their account people use the following two criteria to solve the problem: First, the maximization criterion is to connect as many dots as possible with each line. Second, the progress-monitoring criterion is to determine the ratio between the remaining strokes and dots after each move. For example, after connecting three dots with the first stroke, there are still three strokes left to connect the remaining six dots. Evaluating these criteria, a person with an exceptionally high look-ahead may immediately recognize that the problem is unsolvable within the boundaries of the 3 × 3 dot matrix. The reason is that no configuration of three strokes will satisfy the criteria if one stays within the matrix. This, in turn, may then trigger the search for a promising state such as the non-dot point extension outside the matrix that is required to solve the problem.

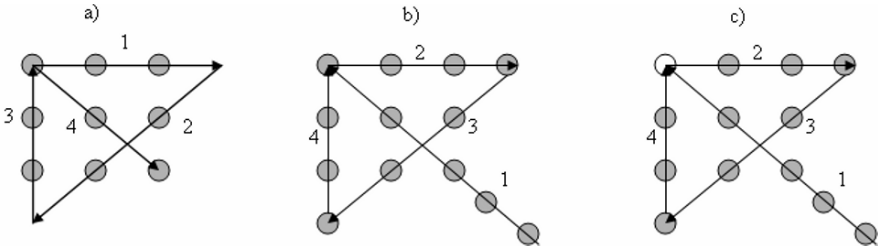


Fig. 4. a) The 9-dot problem and its solution; numbers indicate the sequence of moves. b) The 13-dot variant and its solution. c) The 12-dot variant and its solution.

To test their assumptions empirically, MacGregor and colleagues used several variants of the nine-dot problem. Using the 13-dot problem (Fig. 4b) they demonstrated that problem solvers tended to select such moves that connect as many dots as possible (verifying the maximization criterion). Introducing the 12-dot problem (Fig. 4c) they had no problem to use non-dot extensions, if they satisfied the progress-monitoring criterion. In further studies, MacGregor and colleagues successfully applied the assumptions of maximization, progress-monitoring, and look-ahead to the eight-coin problem (MacGregor *et al.*, 2001; Ormerod *et al.*, 2002), and to another set of coin problems (Chronicle *et al.*, 2004).

The two accounts discussed in this section emphasize the importance of heuristics for problem solving. Both views can be considered as direct extensions of the classical PST. Insight requires "nothing special" except particular strategies. Thus, heuristics are believed to be the driving force behind problem solving. If the appropriate heuristics are known and applied, then insight problems do not differ from other problems (Chronicle *et al.*, 2004, p. 26):

> "Our view of the processes of solution discovery in insight problem solving indicates links between insight and conventional problem solving, suggesting that accounting for insight lies within the scope of unitary cognitive architectures ... and adaptive control of thought ... ."

7 Representational Change and Insight

Another account that builds on PST, but postulates that there are "special" processes of restructuring, is Ohlsson's (1992) representational change theory (RCT) of insight problem solving. In this theory insight problems are considered to be special because they tend to trick the problem solvers into representing the problem in a way that does not allow them to solve it. In many cases the faulty initial representation needs to undergo a fundamental change before the problem can be solved.

Why do people generate inadequate problem representations? The reason is that the initial encoding of the problem depends on the problem solver's prior knowledge. During encoding problem solvers try to apply seemingly appropriate knowledge to the problem, and this affects how the problem elements are grouped perceptually and conceptually. Prior knowledge also dictates which problem elements are selected and which elements are ignored, and constrains what is considered as a solution for the problem. As a consequence, the initial problem representation can be misleading in many ways. It may not contain the crucial problem elements, the elements may be grouped in the wrong manner, or the goal of problem solving may be too narrow in order to allow a solution.

If problem solvers have the wrong problem representation they will sooner or later hit an impasse, a state of mind where problem-solving behavior ceases and where they do not know what to do next. A representational change becomes necessary before new solution paths can be generated. Depending on what is wrong with the initial problem representation, different perceptual and memory processes can lead to a representational change. It is assumed that these processes operate outside of consciousness (Schooler *et al.*, 1993) and thus create the impression of a sudden insight once the representation has changed and new solution paths become available.

The assumptions of RCT have been tested using simple matchstick arithmetic tasks (Knoblich *et al.*, 1999; Knoblich *et al.*, 2001; Knoblich *et al.*, 2005; Knoblich and Öllinger, 2006; Öllinger *et al.*, 2006; Öllinger *et al.*, 2008). Matchstick arithmetic tasks present the problem solver with an equation consisting of numerals and the arithmetic operators +, −, and =. The task is to move a single matchstick in order to generate a true (correct) expression.

Problem type		Initial state	Chunk description	Solution	Problem difficulty
formal	Base	$X = Y - Z$	loose chunks	$X = Y - Z$ (true)	+
example		**VIII=IV+VI**	**VIII=IV+VI**	**VIII=IV+IV**	
formal	Chunk dec.	$X = Y - Z$	tight chunks	$X = Y - Z$ (true)	++
example		**VI=VI+V**	**VI=VI+V**	**XI=VI+V**	

Table 1. Problem types of the matchstick arithmetic task that require a change of the goal representation: chunk decomposition.

Two characteristics of these problems are worth noting. The first is that single matchsticks are perceptually grouped to form meaningful chunks. For example, in the problem displayed in the first row of Tab. 1 the two slanted matchsticks are automatically grouped to form the letter **V**. Similarly, +, =, and so on are immediately recognized as “meaningful chunks”. There are two types of chunks. *Loose chunks* are meaningful entities that can be easily decomposed, for instance **VI** into **V** and **I**. The single Roman numeral **I** can be easily moved within the equation because it is meaningful in itself. By contrast, a *tight chunk* consists of constituents that have no meaning by themselves. For instance, decomposing **X** results in two slanted sticks that have no meaning within the context of the task.

Applying RCT one can predict that problem solvers will initially treat tight chunks as being non-decomposable. That is, the single units forming a tight chunk will initially not be represented separately. For some problems this is an inadequate problem representation because they require transforming the Roman numeral **V** into the Roman numeral **X**. This can be achieved by moving one match, but only if the tight chunk **X** is decomposed into its constituents **** and **/**. We call the process that triggers this change *chunk decomposition.*

A second characteristic of the problems worth noting is that they are likely to activate the problem solver’s knowledge of simple arithmetic. This knowledge will lead to an initial goal representation that constrains the space of possible solutions that problem solvers will potentially consider. The goal representation determines which operations can be applied to the encoded problem elements and guides the problem solving process. Imposing such constraints is an automatic and unconscious process that often helps to reduce the problem space. However, these constraints can also narrow the problem space to an extent that a solution becomes impossible.

Problem type		Initial state	Constraint relaxation	Solution	Problem difficulty
formal	Base	$X = Y - Z$	—	$X = Y - Z$ (true)	+
example		**VIII=IV+VI**	—	**VIII=IV+IV**	
formal	Operator	$X = Y - Z$	$X\mathrm{Op}_1 Y\mathrm{Op}_2 Z$	$X\mathrm{Op}_1 Y\mathrm{Op}_2 Z$	++
example		**IX=VI-III**	**IX=VI-III**	**IX-VI=III**	
formal	Tautology	$X = Y - Z$	$X\mathrm{Op}_1 Y\mathrm{Op}_2 Z$ and $\mathrm{Op}_1 = \mathrm{Op}_2$	$X\mathrm{Op}_1 X\mathrm{Op}_1 X$	+++
example		**VI=VI+VI**	**VI=VI+VI**	**VI=VI=VI**	

Table 2. Problem types of the matchstick arithmetic task that require a change of the goal representation: constraint relaxation.

Consider the following example. Prior knowledge of simple arithmetic suggests that values change and operators do not change (see Tab. 2). Therefore, it is likely that problem solvers form an initial goal representation that represents values as variable and operators as constant (Var1 = Var2 + Var3). For some problems this representation is inappropriate to obtain a successful solution. Ohlsson (1992) postulated that a second unconscious restructuring process relaxes the self-imposed constraints on the goal of problem solving when problem solvers hit an impasse. This process is called *constraint relaxation.* It generates a more flexible goal representation. For instance, representing arithmetic operators as variable activates moves that manipulate the operators. The tautology type in Tab. 2 illustrates that certain tasks require the relaxation of two or more constraints. In addition to conceiving the operator as variable, a second constraint – equations consist of two different kinds of operators – has to be overcome. Only in this case one can conceive of solutions where both operators are equal signs.

Knoblich *et al.* (1999) provided empirical evidence supporting the assumptions outlined above. In several experiments they asked participants to solve different types of matchstick arithmetic tasks that differed in their need to decompose tight chunks and to relax initial constraints on the goal. The results show that chunk decomposition and constraint relaxation are two independent processes that can lead to representational change. In particular, problems requiring the decomposition of tight chunks are much more difficult than problems that only require the decomposition of loose chunks. Moreover, the problem difficulty increases dramatically with the number of self-imposed constraints that must be relaxed to solve a particular problem. The tautology type is more difficult than the operator type, and the operator type is more difficult than the base type (see Tab. 2; Knoblich *et al.*, 1999; Knoblich *et al.*, 2001; Öllinger *et al.*, 2006; Öllinger *et al.*, 2008).

Recently, Öllinger and colleagues (2008) raised the question whether set effects (Luchins, 1942) can be interpreted as resulting from representational change. Their basic idea was that the repeated solution of similar problems induces a gradual representational change which continuously narrows the space of solutions considered. Imagine you are asked to solve a variety of matchstick problems which, conforming to your prior knowledge of arithmetic, can be solved by manipulating values. Will this increase the difficulty of problems that require constraint relaxation (manipulating operators) even more? According to Luchins this should be the case because the same solution procedure was used repeatedly when solving problems that require manipulating values.

Öllinger and colleagues found that this is not the case. Repeatedly solving value problems does not increase the difficulty of problems that require constraint relaxation. The reason is that they afford an initial goal representation consistent with one's prior knowledge of arithmetic that is already dominant and cannot be narrowed further. However, if participants were asked to solve a number of problems that require manipulating operators, the solution of a subsequent problem that requires manipulating values was strongly impaired. In this case, the repeated manipulation of operators biased the goal representations towards representing the operators as variable and the values as constant. As a consequence it became more difficult to solve problems that require a manipulation of values. This demonstrates that a new insight, repeatedly applied, can overwrite existing prior knowledge.

Another interesting finding is that having one insight is likely to reduce the likelihood of having another insight. If people solved problems that repeatedly required the decomposition of tight chunks (**X** → **/** → **V**) the solution of problems that required producing a tautology became almost impossible: Chunk decomposition (one insight) reduced the likelihood of constraint relaxation (another insight). This provides further evidence that chunk decomposition and constraint relaxation are two different processes. Chunk decomposition pertains to perceptual aspects of the problem element whereas constraint relaxation pertains to the solutions a problem solver can conceive of conceptually. Problem solvers found it difficult to switch from being flexible with regard to the percept to being flexible with regard to conceptualizing the problem in a new way.

Another promising approach to address insight empirically is recording people's eye movements while they solve problems (Knoblich *et al.*, 2005). Eye movements provide a more fine-grained behavioral measure than solution times or solution rates. Therefore they allow one to test more specific predictions that could not be tested with performance measures. For instance, Knoblich *et al.* (2001) used this technique to test predictions derived from RCT. Participants attempted to solve matchstick arithmetic tasks (base, operator, tautology, and chunk decomposition) that were more or less likely to require a representational change.

Three predictions were tested: First, during an impasse, the problem solving behavior should cease to some extent. Therefore problem solvers should more often stare at a problem without testing particular solution ideas. The results confirmed this prediction. Mean fixation duration increased for problems that required a representational change. This provides evidence for the assumption that people do encounter impasses during insight problem solving.

The second prediction was that the initial goal representation should be biased towards the values, and therefore values should initially receive more attention (eye movements) than operators. Indeed, participants spent much more time looking at the values than looking at the operators during the initial stages of problem solving.

The third prediction pertained to differences between successful and unsuccessful problem solvers. Successful solvers of insight problems should gradually spend more time looking at the crucial problem elements than unsuccessful problem solvers. It was found that in later stages of problem solving successful problem solvers gazed longer on the operators and the critical tight chunk.

These results support the concepts of impasse and representational change. The problem representation determines which parts of a task problem solvers attend to. Successful problem solvers differ from unsuccessful problem solvers in their ability to shift their attention to previously neglected parts of the problem. Knoblich *et al.* (2001) showed that the shift of attention likely results from a preceding change in the problem representation.

Grant and Spivey (2003) addressed the complementary question whether an externally triggered shift in attention to a crucial problem element could affect the solver's problem representation. They performed two experiments in which they asked participants to solve Duncker's tumor problem (Duncker, 1945): "Given a human being with an inoperable stomach tumor, and lasers which destroy organic tissue at sufficient intensity, how can one cure the person with these lasers and, at the same time, avoid harming the healthy tissue that surrounds the tumor ?" The solution is to use two lasers radiating at the tumor from different angles, so that their beams meet at the location of the tumor. The addition of the intensities of the beams provides the necessary energy to destroy the tumor, while the reduced intensity of the single lasers leaves the surrounding tissue unharmed.

Grant and Spivey provided a simple schematic drawing the problem solvers were looking at while attempting to solve the problem. The tumor was simply depicted as a small solid oval, with a circumscribing oval representing the skin. In a first experiment they found that successful problem solvers looked significantly longer at the skin than unsuccessful problem solvers who looked longer at the tumor. In the second experiment, they tested whether drawing a problem solver's attention to the skin would increase the solution rates. They introduced three conditions. In the first condition the skin pulsated slightly, in the second condition the tumor pulsated slightly, and in the control condition they presented a static display. They idea was that participants' attention was attracted by the pulsating portion of the display. In line with their hypotheses they found that solution rates were significantly increased in the pulsating skin condition compared to the other conditions. This result is quite astonishing because this simple manipulation was much more effective in increasing the solution rates than a variety of explicit verbal hints that had been tried in previous research.

Although RCT can quite well explain why people encounter impasses and which processes can help to resolve these impasses, it is less successful in explaining what happens before and after an impasse. Furthermore there is growing evidence that insight problems often have multiple sources of difficulty that need to be disentangled, some related to heuristics, some related to representational change. This has led to attempts to compare and integrate the two types of accounts which we will discuss in the next section.

8 Heuristics, Representational Change, and Insight

Jones (2003) contrasted the predictions of CSPT and RCT (see also Knoblich *et al.*, 2005). He asked participants to solve problems taken from the car park game while he tracked solvers' eye movements. In this game, one needs to maneuvre a taxi car out of a car park with other cars blocking the exit way. In particular, one needs to figure out how to clear the exit way so that the taxi can leave the car park. In some problems, the taxi itself needs to be moved back and forth before an exit way has been created. Jones (2003) suggested that these problems require insight because problem solvers impose the constraint that the taxi can only be moved after an exit way has been created. Accordingly, in line with the assumptions of RCT, he expected for such problems that problem solvers encounter impasses.

Furthermore, he tested the assumption of CSPT by determining problem solvers look-ahead value. He expected that problem solvers having a greater look-ahead value (see above) should be more successful in solving insight problems, because they should encounter impasses earlier than problem solvers with a smaller look-ahead value (cf. Sec. 6). A higher look-ahead value should be a predictor for a successful solution.

From the eye movements it was determined whether problem solvers encountered impasses before they carried out the crucial taxi move, and how many moves they could plan ahead. Jones found that all participants who successfully solved the problem encountered one or more impasses before they moved the taxi for the first time, and that participants with a greater look-ahead completed the problem significantly faster and needed significantly fewer moves than participants with a smaller look-ahead value.

These results led Jones (2003) to propose that insight problem solving can be best understood if one integrates CSPT and RCT rather than treating them as competing explanations (Jones, 2003, p. 1026):

> "The dynamical constraint theory essentially covers insight up to the point at which insight is sought. [...] The representational change theory on the

> other hand covers how insight will be achieved, and, therefore, the point at which insight is sought is the beginning point of the theory."

Öllinger *et al.* (2006) further investigated the interplay between heuristics and representational change. They modified the matchstick arithmetic and added to each problem type (see Table 2) an additional value move. Thus the new tasks required two moves for a successful solution. Moreover, they created two sets of problems. The first set consisted of problems that required an additional value transformation reducing the distance to the goal. The second set consisted of problems that required an additional value transformation that initially increased the distance to the goal. Distance was defined as the numerical difference between the left-hand side and right-hand side of the equation. For instance, the equation **VI** = **IV** + **VI** has a distance of four. Solving the problem requires two steps, first to apply a value move that changes the **IV** into a **VI**, thus increasing the distance to six. The solution is a tautological structure **VI** = **VI** = **VI** with distance zero.

This task modification allowed Öllinger and colleagues to test assumptions of CSPT and RCT. According to CSPT, the distance measure is nothing else than a maximization criterion. Reducing the distance makes the right and left side of the equation more similar (hill climbing). The criterion for progress is based on assessing, after each move, whether there is an available consecutive move that can equalize the left and right side of the equation. CSPT predicts that tasks requiring two moves that reduce the distance between the left and the right side of the equation should be easier than tasks requiring a move that increases the distance. RCT predicts that the problem difficulty is driven by the degree of representational change required. That is, problems requiring a value move plus a move that produces a tautological structure should be significantly more difficult than problems requiring a combination of two value moves.

Öllinger *et al.* (2006) found that the problem difficulty varied according to whether or not a representational change was required. Problem difficulty was not influenced by the kind of value move. It did not matter whether the moves increased or reduced the distance. Thus, it seems that problem solvers did not apply the maximization criterion of CSPT. In addition, the outcomes indicated that two-move problems were much more difficult than one-move problems (Knoblich *et al.*, 1999; Knoblich *et al.*, 2001; Öllinger *et al.*, 2006; Öllinger *et al.*, 2008). This shows that the larger problem space in two-move matchstick arithmetic tasks was an additional source of problem difficulty. Öllinger and colleagues suggest that the main source of difficulty in insight problems is the necessity to change the problem representation. Heuristics sometimes help to reduce large problem spaces and may therefore reduce the time a problem solver spends exploring fruitless solution paths.

A recent study of the nine-dot problem (Fig. 4a) by Kershaw and Ohlsson (2004) further underlines that the number of sources of difficulty a problem poses must not be underestimated. The classical explanation for the high problem difficulty of the nine-dot problem is that problem solvers initially do not consider moves that go beyond the virtual square formed by the dots (Ohlsson, 1992; Scheerer, 1963). Accordingly, the insight needed for the solution is to realize that the lines can be extended to non-dotted locations outside the virtual square – that is, a representational change (draw beyond the barriers) can solve the problem. However, Weisberg and Alba (1981) showed that people did not benefit from hints that told them to relax this constraint. As described above, MacGregor *et al.* (2001) claimed that the main source of difficulty is the application of the appropriate heuristics.

Kershaw and Ohlsson compiled the contradictory evidence on the nine-dot problem and concluded that it entails four sources of difficulty. First, in line with the classical account an essential amount of the problem difficulty is the necessity of drawing beyond the virtual boundaries. The second source of difficulty is the shape of the solution. The configuration of lines required is quite extraordinary and therefore both hard to find and hard to apply. The third source of difficulty is the size of the solution space: The four consecutive moves create a large problem space and moving beyond the virtual boundaries considerably increases this problem space. Finally, using variants of the standard nine-dot problem, they found that it is difficult to change direction at locations that do not contain dots. They concluded that insight problems often have a number of sources of problem difficulty that require the contribution of different cognitive processes. This indicates the necessity to construct problems for insight research that permit systematic variations of particular sources of difficulty.

In the next section we will give a short review of current findings on neural correlates of insight problem solving. Although this research has been caried out for less than a decade, it has already provided some interesting results that can help to improve functional explanations of insight.

9 Neural Correlates of Insight Problem Solving

Luo and Nicki (2003) performed the first functional magnetic resonance imaging (fMRI) study on insight in order to determine whether particular brain regions are activated in insight problem solving. They asked participants to solve Japanese riddles that either require a reinterpretation of the concepts involved or not. A typical riddle is like this: "What is the thing that can move heavy logs but cannot move a small nail?" The answer is a river. In a first step a number of participants were shown the solution to the riddles and asked whether the solution was surprising. Those riddles that had a surprising solution were used in the "insight" condition and those that were not surprising were used in the "non-insight" condition.

The fMRI technique allows one to infer from magnetic activations in a particular brain site how much oxygen the blood stream currently transports in this region. It is postulated that large oxygen consumption is an indicator for involvement of a particular brain area in a particular task (hemodynamic response). By contrasting the blood flow between insight and non-insight riddles they found that the right hippocampal system was more strongly activated when solutions of insightful riddles were presented to the problem solvers.

The hippocampus is suspected to be the gateway to the long-term memory system. That is, this structure might be responsible for encoding and addressing new information and conveying it, after a delay, to long-term memory (McClelland *et al.*, 1995). Luo and Nicki provided three possible roles the hippocampal system might play. First, the stronger activation for insightful solutions could be due to the formation of novel associations among already existing concepts. Second, the hippocampus might be involved in breaking unwarranted mental fixation. Third, because the hippocampus plays an important role in spatial-orientation tasks it is conceivable that the stronger activation in the insight condition reflects the formation of a new reference frame.

The common denominator of all three possibilities is the involvement of the hippocampus in building or permitting “something new” which would imply an important role of the hippocampal system in representational change. A further study emphasizing the importance of the hippocampal system in insight was conducted by Wagner *et al.* (2004). Participants were presented with digit strings and were asked to apply two rules to these sequences. However, all sequences could be solved according to a third, hidden rule that greatly reduced the difficulty of the task.

Wagner and colleagues investigated whether the likelihood that participants discovered the hidden rule increases after sleep. They pointed out that such strategy changes are very similar to insight insofar as they happen very suddenly and without any recognizable effort on the part of the person. After a long training phase participants in one condition slept for 8 hours, the other participants stayed awake and waited for 8 hours to continue with the task. There were further conditions controlling for the effects of fatigue. Surprisingly, it was found that the group that had slept detected the hidden rule much more often than people who had not slept for the same time. Wagner and colleagues explained the finding as a consequence of consolidation and restructuring new memory representation during nocturnal sleep. Converging with Niki and Luo’s (2003) results, the hippocampus was suspected as the crucial region where restructuring takes place.

A further fMRI study (Jung-Beeman *et al.*, 2004) revealed that the hippocampus may not be the only region involved in restructuring. In the experiments by Jung-Beeman and colleagues participants solved a number of remote association tasks. Three words are presented and the task is to find a target word that in combination with the given words results in a meaningful new word or phrase (e.g., given the words pine, crab, and sauce, the target word is apple). After finding the solution they indicated whether or not it was accompanied by an aha-experience. The responses were classified into insight and non-insight solutions and the hemodynamic activations of both conditions were contrasted. The results showed that insightful solutions were accompanied by activation in the right anterior superior temporal gyrus relative to non-insight problems. Jung-Beeman and colleagues argued that this brain region is putatively responsible for linking mental concepts in a novel way and may foster representational change.

There is also evidence from EEG studies that insight problem solving acquires other neural resources than the solution of conventional or non-insight problems. EEGs are recordings of cortical electrical activity. Although their spatial resolution is fairly poor, the temporal resolution of EEGs is very high. A study of event-related potentials (ERP; the averaged EEG signal triggered by a particular event), conducted by Lavric *et al.* (2000), compared the activation patterns between tasks requiring either analytic reasoning (the Wason selection task) or creative problem solving (Duncker's candle problem, see above). Furthermore, participants were asked to count simultaneously auditory stimuli – the events that triggered the onset of the ERP signal.

Lavric and colleagues predicted that counting would disturb analytic reasoning, because it recruits the same brain sites, but not creative problem solving. This was what they found. The main result was that two factors in the P300 component could be extracted (one located frontally, the other left-lateralized) that differed between analytic and creative problem solving. Moreover, the P300 was located more frontally during analytic problem solving compared to creative problem solving (see also Lavric *et al.*, 1998; Mai *et al.*, 2004). This suggests that insight involves non-analytic modes of thinking.

Jung-Beeman *et al.* (2004) reported a further EEG study where they found that a sudden burst of high-frequency (gamma-band) neural activity precedes insight solutions by about 300 ms and could therefore be a neural marker of the subjective aha-experience. Mai *et al.* (2004) investigated Chinese riddles that either had an expected solution ("no-aha" condition) or an unexpected solution ("aha" condition). They found that between 250 and 500 msec after the onset of the answer "aha" solutions elicited a more negative ERP signal than "no-aha" solutions. The difference wave was located over the central electrode site (Cz) with a peak latency of N380. They speculated that the anterior cingulate cortex may generate this component and that the N380 may reflect conflict detection in "aha" answers that require overcoming constraints imposed by prior knowledge.

Finally, there is neuropsychological evidence that yet another area is involved in insight problem solving. Reverberi *et al.* (2005) investigated the impact of brain lesions on the solution of matchstick arithmetic tasks. Patients with a frontal lesion and healthy controls solved different types of matchstick arithmetic problems (cf. Tab. 2). Surprisingly, patients with a lesion in the dorsolateral prefrontal cortex (DLPFC) turned out to be more successful than healthy controls in solving the difficult insight problems that require to produce a tautology. Reverberi and colleagues argued that the DLPFC might be the site that constrains the space of possible solutions a problem solver considers. They suggest that a lesion in this area reduces top-down control and therefore increases the likelihood that prior knowledge overly constrains the goal of problem solving (however, at the cost of successful analytic thinking). Therefore, it is conceivable that DLPFC is the brain area where a goal representation is formed that integrates elements of a problem situation and prior knowledge.

10 Conclusions

Cognitive psychologists and cognitive neuroscientists build on early research initiated by Gestalt psychology to unravel the mystery of how solutions to difficult problem sometimes appear out of the blue. They work under the assumption that important scientific discoveries are not qualitatively different from groundbreaking insights that have allowed great scientists like Pauli or Einstein to fundamentally change our understanding of natural laws.

Current research suggests that insight problem solving can be described as a process that passes through three phases, the phase preceding an impasse, the impasse phase and its resolution, and the phase after an impasse (Ohlsson, 1992; Knoblich *et al.*, 1999; Öllinger *et al.*, 2006). During the initial phase a problem representation is established. This representation is perceptually and conceptually constrained by the problem solvers' prior knowledge and their experiences. Problem solvers use heuristics to effectively search for a solution in the space of possible solutions defined by the initial representation. At some point, problem solvers fail to find new solution paths. This happens earlier for problem solvers who have a large look-ahead (MacGregor *et al.*, 2001; Jones, 2003).

Then problem solvers get stuck in an impasse, doing nothing (Knoblich *et al.*, 2001) or trying the same unsuccessful solution paths over and over again (Knoblich *et al.*, 1999). During inactive phases, the activation of the initial problem representation gradually drops (Öllinger *et al.*, 2008) and unconscious perceptual and memory processes start to affect different aspects of the problem representation. Chunk decomposition can lead to a regrouping of perceptual elements and constraint relaxation leads to a more flexible goal representation. It is likely that there are other processes that can also affect the problem representation.

Once the problem representation has been altered new solution paths become available and stepwise problem solving is resumed. Heuristics play an important role before and after an impasse but it is not clear to which extent they can actually trigger representational change. There is an ongoing debate on this issue, and further research will tell whether people can develop strategies to change problem representations.

Research on insight problem solving in cognitive neuroscience has just begun, and it is already clear that no single area is responsible for representational change. One neural mechanism that could be important for representational change is memory consolidation in the hippocampus. Such a

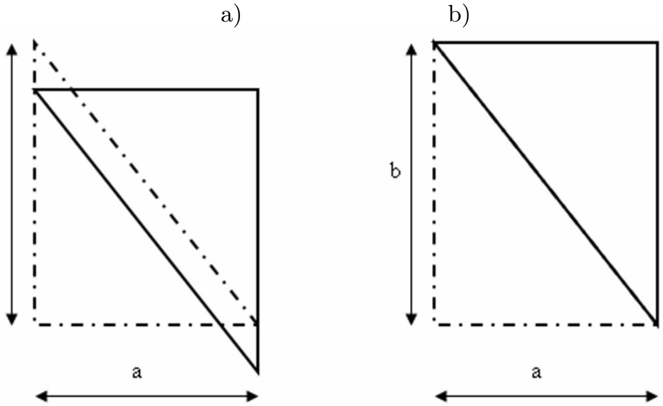


Fig. 5. Solution of Wertheimer's square-parallelogramm problem discussed in Sec. 3: a) restructuring, b) solution.

consolidation could result in conceptual change or in the detection of previously unnoticed regularities. Inferior temporal cortex may also contribute to conceptual change but it is far from clear why the hippocampus is activated in some tasks, why the inferior temporal cortex is activated in others, and whether the results can be generalized at all. Obviously, much more research is needed.

In addition, frontal brain areas seem to be crucially involved in insight problem solving. The anterior cingulate cortex is involved in the detection of conflict and might be involved in making problem solvers realize that they have encountered an impasse. In particular, the repeated failure with an inappropriate representation may be detected and evaluated by the anterior cingulate cortex. Dorsolateral prefrontal cortex seems to be involved in defining the constraints with respect to the goal of problem solving. One of the many issues that will be crucial to address in future research is how different brain areas cooperate in creating new ideas in the problem solver's mind.

References

Ash, M.G. (1998): *Gestalt Psychology in German Culture, 1890-1967: Holism and the Quest for Objectivity.* Cambridge University Press, Cambridge.

Bowden, E.M. and Jung-Beeman, M. (2003): Aha! – Insight experience correlates with solution activation in the right hemisphere. *Psychonomic Bulletin and Review* **10**, 730–737.

Bowden, E.M., Jung-Beeman, M., Fleck, J. and Kounios, J. (2005): New approaches to demystifying insight. *Trends in Cognitive Sciences* **9**, 322–328.

Chronicle, E.P., MacGregor, J.N. and Ormerod, T.C. (2004): What makes an insight problem ? The roles of heuristics, goal conception, and solution recoding in knowledge-lean problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **30**, 14–27.

Dominowski, R.L. and Dallob, P. (1995): Insight and problem solving. In: R.J. Sternberg and J.E. Davidson (eds.), *The Nature of Insight.* MIT Press, Cambridge (MA), pp. 33–62.

Duncker, K. (1935): *Zur Psychologie des produktiven Denkens.* Springer, Berlin.

Duncker, K. (1945): On problem solving. In: J.F. Dashiell (ed.), *Psychological Monographs* **58**(5). The American Psychological Association, Washington (DC).

Grant, E.R. and Spivey, M.J. (2003): Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*, **14**, 462–466.

Greeno, J. (1974): Hobbits and Orcs – Acquisition of a sequential concept. *Cognitive Psychology* **6**, 270–292.

Gruber, H.E. (1995): Insight and affect in the history of science. In: R.J. Sternberg and J.E. Davidson (eds.), *The Nature of Insight.* MIT Press, Cambridge (MA), pp. 397–431.

Jones, G. (2003): Testing two cognitive theories of insight. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **29**, 1017–1027.

Jung-Beeman, M., Bowden, E.M., Haberman, J., Frymiare, J.L., Arambel-Liu, S., Greenblatt, R., Reber, P. and Kounios, J. (2004): Neural activity when people solve verbal problems with insight. *PLoS Biology* **2**, 500–510.

Kaplan, C.A. and Simon, H.A. (1990): In search of insight. *Cognitive Psychology* **22**, 374–419.

Kershaw, T.C. and Ohlsson, S. (2004): Multiple causes of difficulty in insight: The case of the nine-dot problem. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **30**, 3–13.

Knoblich, G., Ohlsson, S., Haider, H. and Rhenius, D. (1999): Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **25**, 1534–1555.

Knoblich, G., Ohlsson, S. and Raney, G. (1999): Resolving impasses in problem solving. In: M. Hahn and S.C. Stoness (eds.), *Proceedings of the 21st Meeting of the Cognitive Science Society.* Erlbaum, Mahwah (NJ), pp. 276–281.

Knoblich, G., Ohlsson, S. and Raney, G.E. (2001): An eye movement study of insight problem solving. *Memory and Cognition*, **29**, 1000–1009.

Knoblich, G. and Öllinger, M. (2006): The eureka moment. *Scientific American Mind*, issue #10, 38–43.

Knoblich, G., Öllinger, M. and Spivey, M.J. (2005): Tracking the eyes to obtain insight into insight problem solving. In: G.D.M. Underwood (ed.), *Cognitve Processes in Eye Guidance.* Oxford University Press, Oxford, pp. 355–375.

Koffka, K. (1935): *Principles of Gestalt Psychology.* Harcourt, Brace and World, New York.

Köhler, W. (1921): *Intelligenzprüfungen an Menschenaffen.* Springer, Berlin.

Lavric, A., Forstmeier, S. and Rippon, G. (2000): Differences in working memory involvement in analytical and creative tasks: An ERP study. *Neuroreport* **11**, 1613–1618.

Lavric, A., Rippon, G. and Forstmeier, S. (1998): ERP studies of working memory in reasoning tasks. *International Journal of Psychophysiology* **30**, 95–271, see p. 145.

Lindsay, P.H. and Norman, D.A. (1981): *Einführung in die Psychologie. Informationsaufnahme und -verarbeitung beim Menschen.* Springer, Berlin.

Lovett, M.C. and Anderson, J.R. (1996): History of success and current context in problem solving: Combined influences on operator selection. *Cognitive Psychology* **31**, 168–217.

Luchins, A.S. (1942): Mechanization in problem solving – the effect of Einstellung. *Psychological Monographs* **54**(6), 1–95.

Luchins, A.S. and Luchins, E.H. (1959): *Rigidity of Behavior: A Variational Approach to the Effect of Einstellung.* University of Oregon Books, Eugene (OR).

Luo, J. and Niki, K. (2003): Function of hippocampus in "insight" of problem solving. *Hippocampus* **13**, 316–323.

MacGregor, J.N., Ormerod, T.C. and Chronicle, E.P. (2001): Information processing and insight: A process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **27**, 176–201.

Mai, X.-Q., Luo, J., Wu, J.-H. and Luo, Y.-J. (2004). "Aha!" effects in a guessing riddle task: An event-related potential study. *Human Brain Mapping* **22**, 261–271.

Maier, N.R.F. (1931): Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology* **12**, 181–194.

McClelland, J.L., McNaughton, B.L. and O'Reilly, R.C. (1995): Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review* **102**, 419–457.

Metcalfe, J. (1986): Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **12**, 288–294.

Metcalfe, J. and Wiebe, D. (1987): Intuition in insight and noninsight problem solving. *Memory and Cognition* **15**, 238–246.

Metzger, W. (1986): *Gestalt-Psychologie. Ausgewählte Werke aus den Jahren 1950 bis 1982.* Waldemar Kramer, Frankfurt.

Newell, A. and Simon, H.A. (1972): *Human Problem Solving.* Prentice Hall, Englewood Cliffs (NJ).

Ohlsson, S. (1992): Information-processing explanations of insight and related phenomena. In: M. Keane and K. Gilhooly (eds.), *Advances in the Psychology of Thinking.* Harvester-Wheatsheaf, London, pp. 1-44.

Öllinger, M., Jones, G. and Knoblich, G. (2006): Heuristics and representational change in two-move matchstick arithmetic tasks. *Advances in Cognitive Psychology* **2**, 239–253.

Öllinger, M., Jones, G. and Knoblich, G. (2008): Investigating the effect of mental set on insight problem solving. *Experimental Psychology* **55**, 270–282.

Ormerod, T.C., MacGregor, J.N. and Chronicle, E P. (2002): Dynamics and constraints in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **28**, 791–799.

Reverberi, C., Toraldo, A., D'Agostini, S. and Skrap, M. (2005): Better without (lateral) frontal cortex ? Insight problems solved by frontal patients. *Brain* **128**, 2882–2890.

Scheerer, M. (1963): Problem-solving. *Scientific American*, **208**(4), 118–128.

Schooler, J.W., Ohlsson, S. and Brooks, K. (1993): Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General* **122**, 166–183.

Thomas, J.C. (1974): An analysis of behavior in the hobbits-orcs problem. *Cognitive Psychology* **6**, 257-269.

Wagner, U., Gais, S., Haider, H., Verleger, R. and Born, J. (2004): Sleep inspires insight. *Nature* **427**, 352–355.

Wegner, D.M. (2002): *The Illusion of Conscious Will.* MIT Press, Cambridge (MA).

Weisberg, R.W. and Alba, J.W. (1982): Problem solving is not like perception: More on Gestalt theory. *Journal of Experimental Psychology: General* **111**, 326–330.

Weisberg, R.W. (1992): Metacognition and insight during problem solving: Comment on Metcalfe. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **18**, 426–431.

Weisberg, R W. (1995): Prolegomena to theories of insight in problem solving: A taxonomy of problems. In: R.J. Sternberg and J.E. Davidson (eds.), *The Nature of Insight.* MIT Press, Cambridge (MA), pp. 157–196.

Wertheimer, M. (1912): Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie* **61**, 161–265.

Wertheimer, M. (1923): Untersuchungen zur Lehre von der Gestalt. II, *Psychologische Forschung* **4**, 301–350.

Wertheimer, M. (1925): *Drei Abhandlungen zur Gestalttheorie.* Verlag der Philosophischen Akademie, Erlangen.

Wertheimer, M. (1959): *Productive Thinking.* Harper, New York.

Wertheimer, Mich. (1985): A Gestalt perspective on computer simulations of cognitive processes. *Computers in Human Behavior* **1**, 19–33.

Wickelgren, W.A. (1974). *How to Solve Problems: Elements of a Theory of Problems and Problem Solving.* Freeman, San Francisco.

Exploring Pauli's (Quantum) Views on Science and Biology

Linda van Speybroeck

Department of Philosophy and Moral Sciences, University of Ghent,
FWO–Flanders, Belgium, `linda.vanspeybroeck@UGent.be`

Summary. Wolfgang Pauli is known as one of the most famous physicists of the 20th century. Next to an intensive treatment of physics, his impressive correspondence with fellow physicists also demonstrates a vivid interest in psychology and biology. Reflections on the mind-brain problem and on topics such as causality and evolutionary theory are readily present. In this paper, some central passages in this correspondence are discussed and linked to more current debates in philosophy of science and philosophy of biology. It is shown how Pauli speculatively explored how evolutionary theory can find inspiration in quantum theory and in its related concept of observer-dependency. Contra Kalervo Laurikainen's interpretation, it is argued that Pauli's criticism remains true to a naturalistic view on science and biology.

1 Introduction

Next to being honored by Albert Einstein and being awarded the Nobel Prize in 1945 for his work on the exclusion principle, Wolfgang Pauli generally is described as one of the most prestigious physicist of the early 20th century.[1] According to Torretti (1999, p. 319), much of this is due to Pauli's introduction of "powerful and very original" assumptions into physical theory. Whereas his academic achievements mainly involved the basic theory of atomic physics, i.e. quantum theory, Pauli showed a clear interest in philosophically oriented topics. This interest, awakened or at least fuelled by his godfather, physicist and positivist philosopher Ernst Mach, took him to delve into the history of philosophical ideas on space, time, matter and reality (Laurikainen, 1988). Especially the mind-brain problem – which forms an aspect of the larger mind-body problem and involves disciplines ranging from biology and psychology

---

[1] Strangely so, d'Espagnat (2006) makes no reference to Pauli in his extensive "On Physics and Philosophy", while this very title hints at Pauli's "Writings on Physics and Philosophy" (Enz and Meyenn, 1994).

to philosophy – attracted the attention of Pauli. This interest was shared with some of his physicist colleagues such as Niels Bohr, Erwin Schrödinger and Werner Heisenberg (Smith, 2005).

Atmanspacher and Primas (2006) concisely review also other and perhaps less obvious aspects of Pauli's interest in biology.[2] In the 1950s, at the advent of an emerging "geneticization of biology and human thought" (cf. Fagot-Largeault *et al.*, 2007), particularly neo-Darwinian evolutionary theory captured Pauli's interest. He speculated about the nature of aspects central to this theory. Especially under attention were the randomness of genetic mutations and the sovereignty of genetic inheritance. More generally, Pauli was interested in the kind of explanation evolutionary theory could offer and what the respective roles of causal and teleological or final explanations are in such a theory. According to Atmanspacher and Primas (2006), Pauli formulated "visionary ideas" about these topics, which today continue to be of value in, for example, bioethics.

In this paper, some of Pauli's ideas relating to biology are further explored by placing them against the background of former interpretations of Pauli's correspondence and of older and more recent debates in philosophy of biology. Against this background, Pauli's thoughts light up as a critical questioning of rational science and speculations about possible alternatives to it. These thoughts are hidden inside an impressive correspondence with colleagues and friends. While this correspondence compares to current email traffic[3] because of the quick follow-up between letters from and to Pauli, it also offers some insight on how Pauli interacted gentleman-like with fellow-thinkers and on Pauli's reasoning itself. In reference to the then present state-of-the-art in science and philosophy, it shows the kind of questions he asked, the kind of answers he sought, the possibilities he investigated and found attractive, the doubts and convictions he had.

In agreement with what Laurikainen (1988, p. vii) states in his "Beyond the Atom. The philosophical thought of Wolfgang Pauli" – which can be considered a remarkable interpretation of Pauli's worldview – this correspondence is more open, more speculative, and less "polished" than academic articles generally allow to be. It presents a work or thinking in progress, which challenges us, readers, to be cautious with our interpretation of it.[4] In this thinking, one constant is so much present that it cannot be ignored, i.e. Pauli's concern with *differences in philosophical perspectives on science.* This major concern revolved around two basic questions in philosophy of science: (i) How does

---

[2] Pauli's father, Wolfgang Josef Pauli, was a physician, but also professor and director of a biological institute at the medical faculty of the University of Vienna (cf. Laurikainen, 1988).

[3] In this regard, I can only express the hope for this correspondence to be digitalized soon, which will make Pauli's work practically more open to research.

[4] I must stress that my own interpretation of Pauli is based on a first acquaintance with Pauli's philosophical ideas as written out in his "Wissenschaflicher Briefwechsel".

science relate to reality ? (ii) How do different scientific disciplines relate to one another ? Let us first explore this further.

2 Philosophical Concerns About Science and Reality

2.1 Which Kind of Reality Are We Observing ?

For those (including myself) without an academic training in physics, quantum theory is an obscure science. By means of mathematics, some experimental and lots of theoretical physics, quantum theory presents a perspective on reality that seems to differ quite a bit from the perspective classical Newtonian physics offers. However, both physical perspectives also differ from the view adopted in modern molecular biology. This complicates any view on science in general.

Newtonian physics sees reality in mechanistic terms of moving material particles of which the trajectory can be traced through contact forces and described by efficient causation in terms of cause and effect. This is thought to lead to unbiased explanations and predictions in terms of universal laws. Ever since the scientific revolution of the 17th century, the concept of objectivity has been strongly related to this mechanistic viewpoint. It holds the idea that neither the observer of the object under study, nor the measurements made to capture aspects of the object, influences the object in any fundamental way. In other words, what is measured is assumed to give a realistic picture of what the object is about. Any influence of the observer on the object is assumed to be non-existent. Hence, the philosophical question of "knowledge" being realistic or ontological (about the world) or idealistic or epistemic (about how we, as humans, construct some world) is decided in favor of the former.

A striking feature of quantum physics is that it allows taking a different stance regarding the place of the observer. Generally speaking, quantum physics turns away from the objectivist stance in favor of a constructivist position. That is, at quantum level, an intimate relation is assumed to exist between observation and object, on grounds that observational acts – i.e. measurements – fundamentally influence how the object appears to the observer. Pauli (1954, p. 286; in Enz and Meyenn, 1994, p. 153) compared this assumption with psychological research:

> "Since the unconscious is not quantitatively measurable, and therefore not capable of mathematical description and since every extension of consciousness ('bringing into consciousness') must by reaction alter the unconscious, we may expect a 'problem of observation' in relation to the unconscious, which, while it presents analogies with that in atomic physics, nevertheless involves considerably greater difficulties."

He also saw an analogy with the alchemist conception in which (Pauli, 1954, p. 286)

> "the deliverance of substance by the man who transforms it . . . is . . . identical with the redeeming transformation of the man through the *opus*, which comes about only 'Deo concedente'."

Einstein objected to quantum physics because of its very assumption of observer-dependency. This kind of physics, Einstein argued, is unable to decide upon what is real or not. In other words, it is unable to decide upon the ontological question. However, scientists like Niels Bohr took quantum physics to be descriptive *tout court*, that is, it is *describing* "a" reality and is all about a "communicable human experience" (d'Espagnat, 2006, p. 23). Pauli himself seems to have explored this constructivist position to the fullest, leading him to state that "es geschieht doch wirklich nur etwas bei einer *Beobachtung*" (see Meyenn, 1993, p. 435). Philosophically speaking, when something is taken to occur only through observation, Pauli sought to formulate an alternative to the objectivist stance in classical physics, a stance that Pauli frequently dubbed "the detached observer" (cf. Pauli, 1957).

With regard to these two broad philosophical schemes of thought, modern biological sciences seem to move somewhere in between. Depending on the specific discipline focused on, biology has been labeled both objective (cf. genetics) and descriptive (cf. evolutionary theory). In current molecular biology both stances even seem to be present simultaneously, leading to a hybrid philosophy in which the observer is acknowledged to play a role in how the object is or can be perceived.

More specifically, it is acknowledged that the observing scientist *cuts out* – on the basis of kinds of research questions asked, methods used and conclusions drawn – aspects or parts of a biological system. As such, some factors are drawn to the foreground, while others are not included in the study. Experimental replications here play a tremendously important role in that they help to rule out the interference of unknown factors or observational biases in research. Through an iteration of experiments, subjective observational influences are set within boundaries. They become, literally "in practice", part of a controlled experiment.

The resulting experimental data or explanations rarely ever function in themselves, but are placed into a larger *theoretical model* of the system in question. This activity of constructing coherent models then solves the debate in molecular biology on what is real. Models represent a *description* of reality, of which the *adequacy* can be tested experimentally. In sum, although one can never claim that the model mirrors reality in a one-to-one correspondence, a continued and iterated experimental follow-up of these models allows to make inferences about the *boundaries* within which reality resides and within which observations are influential or not. This allows making the difference between what is possible in theory and what is likely to be real. Biological sciences, then, are not just a matter of observation, but also a matter of reality offering resistance to this observation. This position lies in between a complete abstraction of observational influences and a total surrender to them.

2.2 The Growing Importance of Context and Complexity in Molecular Biology

This hybrid perspective on science has not always been present in molecular biology. In the 1950s – when Pauli developed his interest in biology, as witnessed by his correspondence with Max Delbrück, Walter Elsasser, and others – the mainstream idea in molecular biology was that in DNA it had finally uncovered the true hereditary unit of life. The study of DNA would provide insight in the material nature of life, supporting the conviction that, like physics, also biological theory would have its universal laws and dogmas. This climate welcomed Francis Crick's "central dogma of molecular biology" (cf. Crick, 1958; 1970) and nurtured a stubborn genetic reductionism, which became the dominant view until late in the 20th century. In the 1990s, the conviction that an objective gene-centered perspective was key to understanding biological objects even led to ambitiously set-up genome projects. These were designed to uncover the full DNA sequence of an organism and, with it, the essence of this organism. *In extremis*, genetic reductionism held it possible to trace all phenotypic characteristics of an organism to the genotypic level. In other words, the phenotype of life would be reduced to its "real" ontic status, i.e. the genetic state.

Throughout this gene-centered era, a critical undercurrent continued to defend biological complexity and the importance of taking biological variety seriously instead of abstracting it away. During the 1990s and onwards, this undercurrent mainly attacked gene-centered thinking and worked towards a paradigmatic shift in biology, i.e. from "it's all in the genes" towards "it's not (at) all in the genes". This has been discussed repeatedly in biology (cf. Lewin, 1998; Lewontin, 2001) as well as in philosophy of biology (cf. Keller, 1995; Sarkar, 1996; Oyama *et al.*, 2001).

Today, many aspects of this discussion have become part of mainstream thinking. This is mainly because, although the genome projects did provide a read-out of the full DNA sequence of diverse model organisms, it did not deliver the expected manual to understand the "book of life".

The reasons hereto are that, for starters, genes do not stand in a one-to-one relation with a phenotypic character, but take part in large networks of interactions. Also, gene regulation proves to be extremely complicated by transcriptional and post-transcriptional processes. In these processes, also non-genetic and environmental factors play a role. Finally, it is not a simple matter to find out what a gene is in the first place. Indeed, where before the idea existed that genes offer an explanation, now the idea is that one must explain why, when and under which circumstances a specific stretch of DNA can be labeled a gene in the first place. As a concept, the gene thus has shifted from *explanans* to *explanandum* (Speybroeck, 2000). Knowing about DNA sequences thus barely suffices to say much about a living organism.

Although the DNA level remains crucial in any biological approach towards living organisms, ever since the genome projects the focus is shifting from “DNA only” to a post genomic approach in which, next to genes, also their relations, the complexities of regulatory processes, and the influence of non-genetic and environmental factors are taken into account. With the realization of the limits of a gene-centered perspective on biological organisms, new challenges have come to the foreground. Because biological variety is extremely diverse, playing at diverse levels, it is far from easy – let alone possible – to take all of it into account in one biological model. Although for some the latter remains the ultimate goal for biology (cf. Kitano, 2001), currently the exercise to be constantly aware of what is abstracted away or is accounted for in any model forms a major challenge for practicing biologists. Models *par excellence* hence are the ones (i) abstracting away as much as possible without loosing track of the phenomenon under study, (ii) bringing the distilled factors that do play a role into a functional relation with each other, and (iii) keeping enough openness to include new factors to fine-tune the study of the phenomenon. This kind of model construction is present in, for example, epigenetics and systems biology.

2.2.1 Epigenetics as Uncovering Contexts of Influence

This rationale of partial and continually evolving models is clearly present in the study of molecular epigenetics. This study also offers a naturalistic alternative to a gene-centered approach to biology. Based on the work of developmental biologist Conrad H. Waddington – another fascinating scientist of the 1950s – epigenetics (as the word literally states) takes into account what goes *above* or *beyond* the genetic level (Speybroeck, 2002). While recognizing that what a gene is relates to how it (spatially and temporally) functions in the organism in which it resides, epigenetic research aims to uncover the different contexts that make this gene function possible. These contexts involve biochemical influences or fluctuation which may play within or across cellular and/or organismic “borders” (Speybroeck, 2000).

While in classical genetics the conviction existed that all phenotypic variation can be explained by genetic variations, the epigenetic approach taught us that this is not always the case (Jablonka and Lamb, 1995; Holliday, 1996). More and more examples are found of phenotypic variation that cannot be traced back to changes in DNA, even if this variation is heritable. Consequently, while the DNA level remains an important factor in understanding this variation, it also appears important to take all kinds of regulatory factors into account which set the activity rate of genes within limits, hereby codetermining the roles such genes can play in the complex biochemical networks which make up a living system.

An epigenetic approach towards living organization thus no longer seeks to reduce life to its so-called “essential” genetic level. Rather, it has a contextual view on genes. That is, only when genes are put within their proper regulatory context, their functionality can be observed. Epigenetic researchers then necessarily have to work with an open view on the phenomena they study, looking farther than the genetic level alone (see also Vijver *et al.*, 2002; Speybroeck *et al.*, 2007).

2.2.2 Systems Biology's Modern Holism

Systems biology, an interdisciplinary project that aims at the integration of genomics with other "omic" databases such as proteomics, interactomics and metabolomics (Vidal, 2001), can be seen as the latest endeavor in biology to build such flexible models. The route towards them, however, is long and, indeed, complicated because it not only involves theoretical, but also and mainly practical – i.e. technological – obstacles. One not only has to take into account the complex material and process-like nature of biological systems, but also the kind of observational data current technologies can and cannot give rise to.

This is specifically felt in systems biology because of its reliance on innovating and quickly developing hightech research tools. For example, current high throughput TAP-technology[5] extracts stable protein complexes from cell suspension cultures (Leene *et al.*, 2007). However, the procedure of crushing these cultures before extraction of proteins may give rise to a bias in the resulting complexes because new complexes may have formed after (and thus because) the intracellular structure was crushed. It also remains unclear when the found complexes play a role during cell cycle, because the TAPs are performed on cells in different stadia of cell cycle. These technological and practical limitations are much discussed, showing an awareness of how system-biological models rely on the degree in which the (often biochemical) details about different molecules and their distributions and relations in space and time, and about inputs to the system that do not behave as regular patterns can be quantitatively measured or – at least – qualitatively taken into account.

This is what former systems approaches towards living systems generally lacked (cf. Vijver *et al.*, 2003). Such approaches, referring to the work of both Ludwig von Bertalanffi and Stuart Kauffman alike, were mainly theoretical and abstract. Whereas they equally promoted the *idea* of biological complexity, their main goal was to find a handful of universal laws by which this complexity could be understood – or, said differently, laws to which biological

---

[5] TAP-technology stands for a technology based on Tandem Affinity Purification, and involves a technical procedure, which extracts protein complexes out of crushed cell cultures. It aims to identify present stable protein complexes, which is an important factor in the characterization of protein function. The term "high throughput" indicates that all proteins present in the sample are screened at the same time. This high throughput methodology currently and quickly infiltrates biological research. Instead of screening the presence of, for example, one mRNA per time, it allows to screen at once the presence of all mRNAs in a sample. This new step in technology also invites to talk about systems biology as a "holistic" science, because in its analyses it no longer needs to omit the behavior of other molecules when focusing on one kind in particular.

complexity could be reduced. The underlying conviction was that biological "details" or "irregularities" are negligible.

In today's molecular biology, it is more and more acknowledged that these details *do* matter, if only because beforehand it is not always evident to determine what is superfluous and what is not for the object under study. One thus might argue that molecular systems biology stands for a *modern form of holism*. This holism is deprived from vitalistic connotations and, more importantly, it does not stand in the way of an analytic approach. As such, instead of throwing away the analytic child with the holistic bathwater, the limits of the analytic approach are tested while proceeding. In pushing the mechanistic-material perspective to its limits, it is experienced that these limits can be stretched and that they are worthwhile of a continued exploration.

In sum, molecular systems biology has started to expand the gene-centered view on organisms. Its approach is a continuing project that stays close to the complicated material nature of biological processes by focusing foremost on cataloguing into databases all kinds of molecular components of a living system. After this phase of "omics biology" and data refinement, a phase of integrative and interdisciplinary biology sets in, in which the combination of and relations between omics data stands central in order to come to a functional interpretation of the data. Methodologically speaking, here the iterative cycle between dry (bioinformatics, mathematics) and wet laboratory experiments is crucial. With it, an interdisciplinary approach towards living organisms is promoted, while mathematics appears as a means to an end, not the end itself. Rather, the "end" is seen in terms of flexible models "open" to the integration of ever more data and new insights.

2.3 From Essentialist Laws to "Paulian Holism"

One can only speculate in how far Pauli would have appreciated the focus in molecular epigenetics on genetic contexts and systems biology's specific call for interdisciplinarity and "analytic-holistic" biology. Aspects of his thinking suggest that he would have been critical towards some elements in these recent developments in molecular biology, while other elements might have aroused some sympathy.

2.3.1 Fundamental Laws and the Role of Variation

The evolution in molecular biology from "gene" to "organism" or, more generally, from "essence" to "context", reminds of another debate in philosophy of science, i.e. the debate on the meaning of physical law.

The general convention in physics is that natural or universal laws exist and are independent of particular conditions or observing parties. This convention determines much of the agenda in classical physics, which amounts to uncovering these so-called fundamental laws. However, in current philosophy of science, as witnessed by Mitchell (2003), the concept of law can be interpreted more flexibly, leaving room to uncover how generalizations function in scientific prediction and explanation (cf. Vijver and Speybroeck, 2006). Mitchell argues how physical laws are not just to be situated on a *continuum* regarding (i) stable conditions and (ii) strength of causal determination, but also regarding (iii) abstractions of contextual matters.

Physical laws are thus like biological "laws" or explanations, necessarily standing within well-defined spaces and times. Depending on the context one chooses, particular regularities and/or variations may not be prominently present and even escape our law-like window of observation. As a result, in both physics and biology not *universal validity*, but rather *universality within contexts* is what matters. If these contexts involve particular conditions or observational influences, these must – under penalty of ending up with inadequate models – feed back to the laws valid in the context of interest.

Did Pauli follow the traditional scheme of abstracting away phenomenological reality until some phenomenological-free stable core or essence is uncovered ? At first sight, some kind of ontological reductionism or essentialism seems present in Pauli's work, for example, when he makes the movement from broken symmetries to fundamental symmetries, with the latter not only suggesting invariance in transformations, but also symbolizing some "ultimate truth". At the same time, however, Pauli acknowledged that physical laws never escape to be human expressions or statements about order relations or causality while, simultaneously, phenomenological reality is a necessary condition in the uncovering of fundamental laws (cf. Atmanspacher and Primas, 2006). This fits the way models are interpreted in molecular biology. Pauli also doubted whether a complete reduction of phenomena to fundamental laws or first principles is possible. While Pauli acknowledged that first principles capture an aspect, indeed an essence, of the phenomena, he also acknowledged that they cannot reconstruct phenomenological variety.

This reevaluation of variety in Pauli's thinking is an important element as it links to biological sciences in which the concept of variety since long has been given a central place. To name but a few, organismic or phenotypic variety was key to Charles Darwin's evolutionary theory and to classical taxonomy, while mutational or genotypic variety was and still is a crucial element in both molecular and population genetic research. Increasingly, molecular research also is taking non-genetic variety into account.

How biological variation nonetheless troubled Pauli will be discussed in Section 3. This section deals with how Pauli looked at biological (specifically evolutionary) science. On a more general level, Pauli's struggle with variation seems to play a role in his view on physics as a maturing science. Inspired by a Platonic view on mathematics and by his reasoning about the symbolic universe humans live in, Pauli saw physics as a maturing science because of its shift from visually intelligible models towards mathematically formulated abstractions beyond sensory perception (cf. Atmanspacher and Primas, 2006).

In biology, approaches taking this very route appeared less successful. Throughout its history, biology had to learn that science can "mature" also in the opposite direction. For example, whereas systems biology demonstrates that biology much depends on developments in mathematics and bioinformatics, these disciplines continually face the challenge to deal with less instead of more abstractions if they want to be of value to biological sciences. Because of the detailed complexity of biological systems and their dependence on both rapidly and slowly changing environmental input, these disciplines also need to develop tools that allow a smooth incorporation of new observational data.

Moreover, the growing complexity of current molecular models urges scientists to develop means to literally visualize this complexity. This kind of "Anschaulichkeit" is not – as Pauli had it – simply a matter of wanting to see things, driven by a childlike urge to see concrete illustrations (Atmanspacher and Primas, 2006), but a matter of taking the constant flux of matter seriously. Living organisms are moving entities, hence visualizing this movement by known parameters that together characterize it is an important step in our understanding of biological systems, acknowledging their dynamicity. As a result, next to the classical diagrams or arrow models, systems biology is not only into mathematical modelling,[6] but also into the integration of mathematical models in *in silico models* or *virtual cells* which are built to simulate the workings of a living cell or modular processes in it (cf. Gershon, 2002).

2.3.2 Holism as an Inseparable Union Between Mind and Matter

One cannot escape the fact that Pauli had an issue with Western scientific thinking, especially with the concepts of objectivity and observer independency. He looked for a "new idea of reality" (German: "neue Wirklichkeitsidee"), i.e. "the idea of the reality of the symbol" (German: "die Idee der Wirklichkeit des Symbols"); see Meyenn (1993, p. 559). This comes down to the idea that our immediate observations of reality are linked to each other

---

[6] Some interpret systems biology as a mathematical endeavor. Leroy Hood, for example, sees biology as an "informational science". He and his colleagues argue that (Ideker *et al.*, 2001, p. 344) "the Human Genome Project has propelled us toward the view that biological systems are fundamentally composed of two types of information: genes, encoding the molecular machines that execute the functions of life, and networks of regulatory interactions, specifying how genes are expressed. All of this information is hierarchical in nature: DNA → mRNA → protein → protein interactions → informational pathways → informational networks → cells → tissues or networks of cells → an organism → populations → ecologies. Of course, other macromolecules and small molecules also participate in these information hierarchies, but the process is driven by genes and interactions between genes and their environments. The central task of systems biology is (a) to comprehensively gather information from each of these distinct levels for individual biological systems and (b) to integrate these data to generate predictive mathematical models of the system". Although they see this goal only fulfilled in a (more or less) distant future, the aim is to bring biology from a descriptive and qualitative science to a predictive and quantitative science. It is doubtable whether Pauli would subscribe to this aim.

through symbols. It is, thus, through symbols that we come to describe reality. But what does this mean?

A "symbolic reality" need not necessarily differ much from the way models function in, for example, systems biology. But while Pauli focused on the role which subjective elements play in "symbolic" models, biological researchers rather focus on how one has to deal with such models in order to learn about reality. The "objective" component gets the major attention. But this is not to say that biologists altogether exclude subjectivity from science. Rather, the attention is directed to where research biases (including biases in measuring or interpreting data) can arise *and* how these can be reduced.

This treatment of subjectivity in modern science probably would not have satisfied Pauli. For example, he can be seen to stress that one is never absolutely sure about the "necessary consequence of the influence, unknown in principle, of the measuring instruments on the observed system" (Pauli, 1948, p. 308). In other words, there *always* will be research biases. Moreover, and partially as a result of such biases, scientific models display *statistical* instead of *absolute* causality and cannot claim to present absolute knowledge. Still, these remarks are well taken in current scientific thinking. Also the idea that on the basis of this statistical causality, "the universe should again be seen as an organism, not a clock" (Laurikainen, 1988, p. xv), is perfectly legitimate and taken at heart in current systems biology whenever organisms are seen as complexes of interacting networks instead of simple genetic recipes. Even more, with the acceptance of statistical causality in biology, prediction is taken to shift from absolute to statistical as well. In other words, biological science is learning to deal with uncertainties.

Yet it seems that Pauli did not follow this reasoning on biological science. At least in Laurikainen's interpretation it is argued how the loss of absolute causality was taken by Pauli as a sign to profoundly distrust rational research and prediction. Instead of exploring the possibilities left for an analytic science, Pauli rather demonstrated a strong sympathy for holistic perspectives on reality. He considered them to be possible alternatives to the detached-observer stance of modern science and pleaded that only in holistic terms an understanding of what lies at the basis of our contemplation of reality can be reached. More specifically, on the basis of a psycho-physical unity, he stressed the upheaval of any distinction *a priori* between mind and matter, religion and science, faith and knowledge.

A straightforward interpretation of the implications of such a holism is difficult to extract from Pauli's writings. As Laurikainen (1988, p. ix, emphasis added) describes it, this is because Pauli's philosophical speculations foremost mean to imply an "emphasis on the *intuitive* comprehension of wholeness instead of the exact detail". Regarding Pauli's thinking, the term "intuition" seems especially well chosen.[7] Following the Merriam-Webster Online Dictio-

---

[7] In a similar vein, Laurikainen states that in order to understand Pauli's criticism on objective science, "an appeal to intuition is needed". It follows that,

nary, intuition refers to "the power or faculty of attaining to direct knowledge or cognition without evident rational thought and inference". This fits Pauli's speculations about the nature of reality, which he assumed literally to be *irrational*, i.e. transcending our rational minds. In Laurikainen's reading, Pauli's view on reality included both an incomplete rational pole and a more "true" or "ultimate" irrational pole. In fact, it reflects Pauli's hope to grasp after all what reality "really" is.

In this sense, Pauli was as much on a quest to find "the truth about reality" as his so-called opponents. The difficulty with Pauli's approach is that, by putting irrationality at the basis of reality, rational tools no longer can be called in to function as means of control. As with quantum wave-particle dualism, one has to accept that reality can be described by mutually exclusive manners, allowing contradictions to play a role. As a result, the door is set wide open to an "anything goes" approach towards reality. In other words, there are no limits to what concepts such as "mind", "spirit", and "reality" can be applied to, inviting metaphysics, but also vitalism and mysticism alike to re-enter modern thinking.

This indeed would signify the end of scientific practice as we know it. But what does it deliver in its place? At best, it delivers some conceptual ideas about how reality might be, but there are no scientific means to further explore these ideas. Hence, much comes down to how any reader interprets these ideas, what one reads into them, or the meaning one adds to them.

As such, and interestingly, Pauli's worldview becomes highly subjective and personal in two directions. Interpreting Pauli is as much about Pauli as it is about our own convictions and beliefs. It allows Laurikainen to portray Pauli – based on his interest in Jungian interpretations of dreams and in the history of alchemist practice – as being almost exclusively in favor of a mystic, almost religious, way of thinking.[8] But it also allows readers, such

---

"in the description of such matters the choice of words is decisive" (Laurikainen, 1988, p. 88). This reminds us to much of philosophy, which – because of a lack of experimental methods – tries to grasp reality through language and conceptual schemes.

[8] Laurikainen also argues that someone like Niels Bohr could not grasp Pauli's idea of holism because of a "strong aversion to mysticism" (Laurikainen, 1988, p. xii). However, this explanation may not be entirely adequate. As Laurikainen describes, Bohr agreed that any phenomenon is a whole in that it always "includes both the object system to be described and the instruments used for the observation [i.e. preparation and registration] of the 'phenomenon'" (Laurikainen, 1988, p. xiii). Instead of crusading against mysticism, it seems more likely that Bohr adhered to a modern holistic thinking as currently present in systems biology: a holism that acknowledges the role of the observer in scientific practice, without preventing that a rational outcome remains possible in terms of models. As we will see, Pauli wanted more than acknowledging the observer – he also wanted to acknowledge single events.

as myself, to doubt[9] that Pauli intended to downplay scientific thinking in favor of *contemplation* or – i.e. referring once again to the Merriam-Webster Online Dictionary – a "concentration on spiritual things as a state of mystical awareness of God's being".

As Laurikainen (1988, p. 20) recalls, Pauli started off his academic career as a positivistic thinker with a strong aversion towards metaphysics. Perhaps the lure of metaphysics kicked in as soon as Pauli experienced that some questions fall out of the scope of objective – and even intersubjective – science or cannot (yet) be answered by it. So some disillusion – or should we say disenchantment ? – in so-called objective science might have featured in Pauli's search for a new idea of reality. Whatever the reason may be, Pauli seemed to have kept a traditional belief in an "ultimate" or "true" approach towards reality. This led him to see objective science as less "true". It also led him to take an active interest in those models of reality that include as much subjective elements as possible. This direction is opposite to current scientific practice.

In a "reverse engineering" style, Pauli explored the implications of including into scientific thinking those aspects we respect most in humankind, i.e. consciousness and free will or indeterminism. Instead of going from matter to mind, Pauli worked his way from mind to matter. This led him to conclude that also non-organic matter should show weak psychic parallel components (see the letter to Fierz of January 19, 1953; in Meyenn, 1999, p. 19). Neither should one exclude that psyche was not present at the onset of duplication of organic molecules in chemical evolution.[10]

However much such conclusions may be attractive, they remain philosophical speculations. In fact, on the basis of quantum principles, the Paulian worldview asks us to take into account that "the universe has a component which cannot be described on the basis of causal analysis but which nevertheless influences the events of the physical world" (Laurikainen, 1988, p. 62). In this kind of thinking, Pauli may have been inspired by Jung's experience in his psychiatric practice with so-called creational acts – events of which the causes are inconceivable (cf. Atmanspacher and Primas, 2006).

However, and here my disagreement with Laurikainen shows, I do not see Pauli going mystical. Rather, I see him trying to give a naturalistic account of what one might understand under such creational acts. It is not about un-

---

[9] Admittedly, my doubts may be due to a limited reading of Pauli's work.

[10] Pauli referred to the neo-Darwinist Rensch: "In particular, Rensch thinks that the 'psychic parallel components' could not possibly have 'suddenly popped up' in an otherwise continuous ontogenesis." (German original: "Insbesondere meint Rensch, die 'psychischen Parallelkomponenten' könnten doch unmöglich in der sonst stetigen Ontogenese 'plötzlich aufgesprungen' sein.") This reference is made in a letter to Fierz, March 5, 1957 (Meyenn, 2005a, pp. 289ff). The presence of weak psychic components is extrapolated to the onset of life. Pauli argued that these components operate microscopically in a "transient phase", after which "causal fixation" sets in.

graspable magic, but rather about natural, hence physical, irregular events or one-time events. Although these may be experienced at the phenomenological level, science lacks the means to measure or quantify them. These events escape the attention of science, because it generally concentrates on regular patterns, and its tools are developed to do so. Experiments only pick up what repeats itself under the same circumstances. What does not repeat itself cannot be captured, but this does not mean it is less real.

I think this is what Pauli saw as an important challenge to science – a challenge that is, at a metalevel, perfectly rational and reasonable. And as far as one can gather from systems biology, time has proven Pauli right on this point. Pauli did not develop an answer on how to handle such phenomena. He distinguished and labeled them as Σ-phenomena: although they are not governed by any regularity, and hence fall out of the scope of what a classical law can describe,[11] they may and do interfere with causally regulated events.[12] This is why Pauli saw quantum physics – in which such Σ-phenomena do pop up – as a perfect tool to complement biological and, especially, evolutionary, theory.

3 Towards a Quantum View on Biology

3.1 Quantum Theory and Deep Questions

Helrich (2007, p. 99) claims that while "physics will continue to refuse a commitment to a specific metaphysics other than that already inherent in science", quantum theory has "compelled physicists to confront deep questions". That is, "if we believe that the quantum theory presents us with *fundamental truths* about the universe, we *must* ask these deep questions – and we find ourselves in the territory of philosophers and theologians" (Helrich, 2007, p. 99, italics added). As pointed out above, while arguing against aspects of rational and objective science, Pauli also continued to think in terms of fundamental truths. Consequently, and at least in his correspondence, he speculated at large about the relation between deep questions and quantum theory. Especially deep questions relating to biology and the status of humankind provided attractive topics for discussion.

---

[11] Laurikainen (1988, p. 55) speaks here of irrational causes, i.e. causes "which cannot be described in the framework of a rational analysis". However, Pauli means to say that these causes are real (i.e. they help to determine which potentially possible point in the state space of a system is actually occupied), but fall out of the scope of our mathematical tools. Laurikainen interprets this as if Pauli means to say that "in this way supernatural things can also be a part of the universe!" (Laurikainen, 1988, p. 55).

[12] Much confusion in reading Pauli's writings may be due to the common-sense interpretation of causality as an empirical concept relating a change in effect to factors preceding that change. This Kantian view is not Pauli's, who rather saw causality as a non-empirical, mathematical category.

Throughout the history of modern mankind, the mechanistic view on life often has been debated. Even for "enlightened" philosophers, such as Immanuel Kant, the main obstacle to go with the mechanistic flow is the paradox that this raises between a deterministic nature on the one hand and free will on the other. Whereas the latter symbolizes a collection of ideas ranging from morality and responsibility to spirituality and divinity, the former is held responsible to behead all these ideas and, with it, to deprive mankind of all kinds of sense-making and of belonging to the world. In this context, quantum physics unintendedly has proven to be a welcome means to revive the concept of free will. In the eyes of some beholders, quantum physics even allows us to construct a complete metaphysical realm. And because this realm is deduced from a respected physical theory, it is at the same time held to be ontologically meaningful or existent.

Schäfer (2006, p. 506), for example, argues that quantum physics "points to transcending aspects of physical reality, and thus of human nature itself, providing new hope that a life with values is not in conflict with our science". This argument generally assumes that, because "elementary particles can exist in states in which they have no definite position in space" (Schäfer, 2006, p. 506, in reference to Heisenberg), the *macroscopic* characteristic of consciousness can be attributed to these elementary particles: "they [elementary particles] display aspects of consciousness in a rudimentary way" (Schäfer, 2006, p. 506). An extrapolation from the macro- to the microscopic level is made.

Unfortunately for the argument, however, it is argued *also* that physical reality at the macroscopic level is "not what it looks like" (Schäfer, 2006, p. 506), and a reduction should be made to reality as it is at the quantum level. This quantum reality is said to transcend other kinds of reality. Only extrapolations from micro- to macrolevel thus seem to be allowed. But then it becomes incomprehensible how the concept of consciousness – a macrolevel concept – should be interpreted at this transcending quantum level; or what it means to claim – in line with idealistic philosophy – that "the background of reality is mindlike" (Schäfer, 2006, p. 507); or how quantum physics leads to the conclusion that "physical reality is part of a divine reality." These problems usually remain unaddressed in the argument.

Also unaddressed is the importance of distinguishing between epistemological and ontological statements. For example, probability waves in quantum mechanics, presenting "just information on numerical relation" (Schäfer, 2006, p. 507), i.e. some sort of statistical value, are taken to be ontological instead of epistemic entities. They are considered a sufficient reason for the existence of free will, because "in processes ruled by probabilities, one can never be sure of the outcome of a specific event" (Schäfer, 2006, p. 510). That quantum systems are "sensitive to gradients of information [i.e. information the observer has]" (Schäfer, 2006, p. 510) similarly is not taken as an epistemic, but as an ontological statement.

In other words, quantum mechanics is not taken as a model or a theory about reality, but as reality itself. Hence it becomes possible to see the quantum level as the founding level, to "argue away" or "philosophically reduce" macrolevel mechanistic features, and to extrapolate to the macrolevel characteristics appropriate at the quantum level – such as "particles can act without any delay on each other, no matter how far apart they are" and "reality is nonlocal, [hence] the nature of the universe is that of an undivided wholeness" (Schäfer, 2006, p. 510). As a result, characteristics that are not experimentally verifiable at the macrolevel are not just "installed" at this level, but seen to supersede experimentally verifiable characteristics. This saves moral free will, but only at the cost of metaphysical modesty.

3.2 Questioning (Evolutionary) Biology

Pauli was convinced that not all phenomenological variation can be deduced from fundamental physics. This bears the risk to end up with a "metaphysical rank growth" less present in Pauli's thinking. Nevertheless, Pauli saw it as a challenge to develop reflections on biology in coherence with the quantum theory he subscribed to. He thought it useful and possible to extrapolate quantum ideas to the biological level. Both basic and specific aspects of quantum theory, but also the acknowledgement of its explanatory value, thus influenced Pauli's speculations on biology. Next, we look into some of the speculations Pauli alluded to in his correspondence. We mainly focus on the letters written between March 1957 and October 1958, the final year of his life.

3.2.1 The Concept of Chance

In a letter of March 1, 1958 (see Meyenn, 2005b, p. 997) to theoretical physicist, George Gamow, Pauli wants "to talk a little bit on *biology*", especially "on the basic assumption of biology in general" with which Pauli admits to "have some difficulties".

His criticism especially touches upon neo-Darwinian evolutionary theory. The whole idea of dubbing it a "theory" implies, for Pauli, the existence of laws applying to the existence and evolution of life. Early neo-Darwinism claims to have found these in terms of random or chance variation at the genetic level and natural selection, i.e. what Jacques Monod has called *chance and necessity.* However, in his letter to Gamow, Pauli doubts (Meyenn, 2005b, p. 997) whether this "orthodox view" of "random mutation (= chance) + natural selection (= chance)" is "sufficient to explain the whole of evolution". Pauli calls evolutionary theory (Meyenn, 2005b, p. 997) a

> "philosophy ... going very far beyond that, which is empirically known. And nobody gives any explanation for the occurrences within a given time of any event, which is important in evolution (as for instance, that a reptile gets feathers)."

According to Pauli, the core of his difficulties with biology exactly resides in the neo-Darwinian use of the term "chance". In evolutionary theory, this term is applied "to single events without connection with probability calculus, in a way that is entirely synonymous with 'miracle' " (Meyenn, 2005b, pp. 997f).

This criticism boils down to Pauli's general remark that quantum theory forces us to give up on the particular, while biology relies heavily on particular events.[13] This very point of criticism also features in the debate on the origin of life. While biologists discuss the specific details of such an origin (whether life started as an RNA-world or as a form of autocatalytic chemical networking), they do not question the idea that life emerged as a one-time event. This bothers Pauli. In his letter to Gamow (Meyenn, 2005b, p. 997f), he comments that when biologists claim that "life has been generated by a chance-combination", they can only mean to say that "nobody knows the causes for this single event".

Pauli's focus on the mathematical side of the problem, i.e. that biology cannot calculate or predict single events in which life originated, leads to the concern that the extremely low probability the emergence of life had equals the improbability of "magical events". What Pauli does not take into consideration is that the origin-of-life debate does not aim to add metaphysical assumptions to biology, whereas the acceptance of magical events would. Magical events are not highly improbable, but impossible. If they are improbable, they should be labeled natural events and at least one plausible naturalistic scenario for their actualization must be possible. So, while biologists today will agree with Pauli's conclusion that nobody knows the exact causes of the origin of life, they will add that not knowing the absolute causes for a single event is not the same as knowing nothing. Research on the origin of life is working towards models that fit current knowledge about living systems and

---

[13] Another classical concept Pauli refers to is "time". Both in classical and quantum physics, time is reversible. This seems in contradiction with macroscopic systems, such as biological organisms. In these systems time appears to be irreversible, i.e. "the direction of time is provided by the increase of entropy in irreversible processes" (Helrich, 2007, p. 105). This paradox stands in the way of a smooth transition from mechanics to thermodynamics, *except* if one gives up the idea of measuring the trajectory of individual molecules. Indeed, entropy is not defined as a molecular property, but as a system property. This also means that in such a scheme, the particular trajectory of a molecule can have "no meaning in our concept of reality" (Helrich, 2007, p. 106). As a result, biological systems should be taken as *systems* of interrelated molecules. Because any measurement on such a system involves an amount of time and volume, this measurement can only present averages. Biological knowledge thus seems to crystallize itself at the ensemble level. This relates to quantum physics in that the concept of single electrons in complex atoms and molecules must be traded for the concept of *probability waves*. The question that Pauli addressed was whether biology faced a similar challenge to give up on the particular in favor of probabilistic or statistical knowledge. For Pauli, however, this thought exercise is related to the unknown influence of the observer, rather than to the influence of extra unknown natural factors that are involved in how a system behaves. In current systems biology, both aspects are taken into account.

insights about the earthly conditions in past times, in order to come to a plausible naturalistic scenario. If such a scenario is possible, Occam's razor makes the need for metaphysical explanations unnecessary. Hence, exit magic.

3.2.2 Efficient or Final Causation ?

For Pauli, the concept of chance not only links to a lack of knowledge about the underlying causes of an event. It also refers to the presumed absence of goal-directedness, finality or teleology in this causation. Evolutionary theory is said to thrive on "blind" or "random" variation. Pauli (in a letter to Fierz of October 8, 1953) speaks of "the *acausal* in natural philosophy", in order to refer to this absence. He claims that[14]

> "it remains still open to what extent the acausal is 'blind' chance, i.e. without *purposeful* meaning (as in quantum mechanics). Maybe we are approaching a new kind of holism in both parapsychology ... and biology. (I think it is implausible that for our attempts to understand biological evolution 'blind' *chance* as a selection factor will suffice – as the neo-Darwinists want it to be. Should not external conditions and mutations (heritable variations of genes) sometimes have formed an *inseparable* (that is, holistic) phenomenon ?) The 'vital' might refer to those two areas (parapsychology and biology)."

In contrast to Laurikainen's interpretation, Pauli again seems to look for a naturalistic holism. He does not seek to refute neo-Darwinism, but rather to refine and/or extend it. In Pauli's view, natural selection based on random variation remains an important factor. The question is whether it counts as a *universal mechanism*, and whether it can *explain* everything one encounters in the biological realm. Pauli doubts this. He does so on the basis that any theory of biological evolution of organisms *at the same time* has to take into account how organisms relate to their environment and to what extent "different kinds of heritable variation" are at play. In other words, is neo-Darwinism based on natural selection and random variation inclusive enough to comprehend the organism *as an organized entity, as a whole ?*

As Atmanspacher and Primas (2006) hint at, Pauli's doubt at least opens up a theoretical space in which to speculate about kinds of heritable variations. Jablonka and Lamb (1995) demonstrated how epigenetics explores this space

---

[14] German original (Meyenn, 1999, p. 283f): "Es ist aber noch offen, wie viel das Akausale ein 'blinder' Zufall, d.h. auch ohne *Zweck*-sinn ist (wie in der Quantenmechanik). Wir nähern uns vielleicht der Erkenntnis einer neuen Art von Ganzheitlichtkeit sowohl in der Parapsychologie ... als auch in der Biologie. (Es ist mir nämlich unplausibel, dass man beim Verstehen der biologischen Evolution mit dem 'blinden' *Zufall* als Auslesefaktor immer durchkommen wird – wie die Neo-Darwinisten es wollen. Sollten dabei nicht äussere Umstände und Mutationen (erbliche Veränderungen der Gene) manchmal ein *unteilbares* (d.h. ganzheitliches) Phänomen gebildet haben ?) Das 'vital' könnte auf diese beiden Gebiete (Parapsychologie *und* Biologie) hinweisen."

in practice. The scientific examples they collected show that the concept of random mutation is only partially correct. On the one hand, the concept correctly captures that research until today does not allow seeing genetic mutations as being adaptive. On the other hand, it incorrectly assumes that each base pare in a genomic sequence of an organism is prone to the same mutation rate. As bacterial research showed, backward and forward mutations do not have the same frequency rate. Genetic hot spots do exist. As research on eukaryotes showed, genomic regions may be more prone to mutational variation because of methylation or other types of genomic imprinting. As research in higher organisms shows, the Weismannian barrier – although only within specific time windows of a developing organism – *can* be crossed. And environmental factors may be found to trigger these kinds of changes.

To stress this shade in thinking (and in research focus!), Jablonka and Lamb pleaded for the use of the concept of "directed mutations". This reminds us of the close and differentiated relation between the environment and the organism (cf. Speybroeck, 2000), and of how at the level of the individual organism this relation seems to direct the evolution of the organism. However, they also stressed that at the molecular level, this "mechanism of directedness" is nothing but the result of plain orthodox natural selection and non-adaptive genetic variation between individuals.

Epigenetics demonstrates that neo-Darwinism in a refined form does not depend on hopeful monsters or miracles as much as assumed by some of its opponents. Instead of clearing the way for adaptive directed mutations[15] and vitalism, Pauli was searching to make such refinements scientifically viable.

---

[15] According to Atmanspacher and Primas (2006), Pauli believed in goal-directed processes and in the existence of causal influences of the environment on inherited properties, thereby passing the Weismannian barrier between cytoplasm and nucleus. They see Pauli's ideas confirmed in current epigenetics, on the basis of which they argue that evolutionary change can happen through both selection and instruction. The latter refers to final causation, a form of causation that is not accepted in biological theory. The authors claim (Atmanspacher and Primas, 2006, pp. 40–41) that "the rejection of final causes does not follow from first principles of physics, but is motivated by our ability to construct causal instruments and machines. The claim that final processes are impossible is a dogmatic metaphysical preconception that should not be accepted uncritically." Atmanspacher and Primas further state that these first principles allow backward and forward processes. These cannot be picked up by the experimental practice in biology, because this practice relies on a belief in the arrow of time (see also Pauli's letter to Gamow in Meyenn (2005b, pp. 997f)). In current biology, only efficient causation presents itself as a plausible view on causality. I agree that the reduction of causation to efficient causation is linked to the technological-analytical approach biology takes towards life. However, as long as alternative viewpoints do not deliver means to work with or provide more adequate knowledge about living systems, they are of limited value. As a second remark, it is important to keep in mind the difference between goal-directed variations and directed or induced genetic variations, as made clear in Jablonka and Lamb (1995, Secs. 1 and 3).

3.2.3 Biological Complexity and Biotonic Laws

The question whether neo-Darwinism is inclusive enough to comprehend the organism as a whole intrigued yet another physicist. Walter Elsasser was fascinated by the huge complexity of living cells, which led him to believe that a fundamental distinction exists between living and non-living entities. This distinction, Elsasser argued, is related to the concepts of causality and determinism. Elsasser assumed that non-living entities can be fully understood in terms of linear causal chains and determinism, while the causal complexity of living organisms makes such an analytic understanding or prediction *in principle* no longer possible.

At the basis of Elsasser's reasoning lies the idea that, at the atomic level, living cells house extremely small energy changes which computational systems cannot pick up. Instead, these signals are abstracted away as mere noise. Also the heterogeneity, out of which the members of any biological class exist, is too complex to be taken into account. All this makes mechanistic models of biological phenomena necessarily at least incomplete. In Elsasser's words (1998, p. 114):

> "While we nowhere deny that it may be possible to 'reduce' such processes [of creative selection] to simple chemical components each of which obeys quantum mechanics, one must always remember that any such simple representation has to be preceded by a selection from an immense reservoir of admissible states".

For this reason, Elsasser thought it necessary for biological sciences to move towards a formal logic of correlations between supra-molecular events. Elsasser here saw an alternative in what he called "biotonic laws", i.e. principles of nature that define regularities not determined by atomic and molecular physics, or derivable from quantum mechanics. In "Reflections on a Theory of Organisms" (Elsasser, 1998), originally published in 1987, Elsasser pointed out four such principles:

(i) Although the number of structural arrangements of atoms in a cell outranges the number of elementary particles in the universe, these arrangements show regularities or patterns at a higher level. This is "ordered heterogeneity".
(ii) Physical laws cannot explain the difference between the many possible patterns and the one pattern *selected* or *realized*, making biology compatible but not uniquely determined by physical laws. This is "creative selection".
(iii) A temporal stability of information, or a "holistic memory", lies at the basis of living organisms and their heredity.
(iv) In living organisms, DNA serves as a material carrier of information. This is needed to make a holistic memory possible.

Whereas DNA stands for an "operative symbolism", more is needed to understand life than the mechanistic operation of the genetic code. Also the whole organism and all processes involved in it must be taken into account. Moreover, an epistemological principle comes in (Elsasser, 1998, p. 114):

> "In the presence of endless complexity the role of what can be known and what must be considered as unknowable has to be checked at every step until some satisfactory order has been achieved. As has been shown so many times in the history of physics, structuration and epistemological reform cannot be separated from each other."

Elsasser saw this as a dualistic feature of life and as a means to overcome reductionistic thinking in molecular biology.

In September 1958, Pauli started a correspondence with Elsasser, after he received and enthusiastically read Elsasser's (1958) book "The Physical Foundation of Biology. An Analytical Study". In a letter of September 12, 1958, to Elsasser, Pauli foremost stated:[16] "I am no biologist either, and am afraid I cannot tell you about the role the 'biotonic' plays in practical experimental biology." This stresses that his reflections are philosophical and theoretical speculations, which may not reach the practical side of biology. Pauli agreed that the concept of biotonic law captures[17]

> "the problem of the existence of biological laws which, on the one hand, do *not* observationally contradict the physical-chemical ones (classically + quantum mechanically), but go beyond these, do not follow from them, on the other hand."

He saw this illustrated in Watson and Crick's model of DNA as a double helix and continued that[18]

> "[Delbrück] thinks it is possible that a *unique assignment* of 'chemical map' and 'genetic map' *is no longer possible in principle*, if the distances in the atom groups concerned become small. ... The 'biotonic' must, if it exists at all, have occurred already in complex macromolecules."

Pauli saw more difficulties in accepting how the concept of "class" is used by Elsasser and in biology in general. In physics, when using (statistical) experiments, one must assume that "all elements of a class may be considered

---

[16] German original (Meyenn, 2005b, p. 1254): "Ich bin auch kein Biologe und kann Ihnen leider nicht sagen, wie es mit dem 'biotonic' in der praktischen experimentellen Biologie zugeht."

[17] German original (Meyenn, 2005b, p. 1253): "[Das Konzept des biotonischen Gesetzes erfasst] das Problem der Existenz biologischer Gesetze, die einerseits observational den physikalisch-chemischen (klassisch + quantenmechanisch) *nicht* widersprechen, andererseits über diese hinausgehen, nicht aus diesen folgen."

[18] German original (Meyenn, 2005b, p. 1253): "[Delbrück] hält es nun für denkbar, daß *eine eineindeutige Zuordnung* von 'chemical map' und 'genetic map' *prinzipiell nicht mehr möglich ist*, wenn die Distanzen der betreffenden Atomgruppen klein werden. ... Das 'Biotonische' muß, wenn es überhaupt existiert, schon immer bei komplexen Makromolekülen auftreten."

alike in these experiments".[19] If this were not the case, Pauli added, there is no reason to repeat the experiment. Nonetheless, biological classes (such as the class of genes) do not contain like individuals. From the perspective of the Laplacian spirit, biology even can be said to consist of "individuals" instead of "classes". In other words: "the Laplacian spirit measures such that for it two horses are as different as a horse and an ox".[20] In order to argue for a stronger concept of class in biology, one has to take into account *both* genetic-biological *and* physico-chemical aspects. But even so, Pauli argued, the problem of the particular sets in principle a limit on "the refinement of measurement" (Meyenn, 2005b, p. 1253). Today, this problem indeed presents one of the greatest challenges for systems biology to overcome. It is up to future technological and conceptual developments to determine how much this problem will continue to weigh.

Elsasser (see Meyenn, 2005b, pp. 1265ff), responding just two weeks later and writing for convenience in English (so he could dictate this letter), devoted his answer to discuss the status of the Laplacian spirit which he saw as "merely a very clever observer". Hence there is no need to add "metaphysical connotations of a demiurg", nor to "stretch the functions of this personage", as Pauli did when claiming that the Laplacian spirit cannot discriminate between classes, but only between individuals. Instead, Elsasser (1958, p. 193) saw this spirit as

> "a mental device that permits to us to overcome ... human inadequacies in our reasoning. It is a construct based on the idea that any interaction can be interpreted as a measurement, the Laplacian spirit being the measurer. Philosophically speaking, this view is equally far removed from that attitude in which atoms and molecules are uncritically treated as perfectly well defined objects similar to billard balls, and from the alternate extreme where objects are considered to exist only when there is a human subject to observe them."

He continued (Elsasser, 1958, p. 209) that this idea

> "gives *concreteness* to the concept of the ideal observer, the Laplacian Spirit. He is now very far indeed from that ethereal being conceived by Laplace and his contemporaries who would measure with infinite precision the positions and velocities of all particles at the beginning of Time. Instead he is merely an *idealization of the interaction of a physcial object with other physical objects*. Such freedom as we may give him refers merely to designing the interacting objects, the measuring instruments, in such a way that their *disturbing effects upon the object to be measured is minimized*, although in quantum theory it cannot be eliminated altogether."

---

[19] German original (Meyenn, 2005b, p. 1253): "... [dass] alle Elemente einer Klasse für diese Experimente als gleich betrachtet werden dürfen."

[20] German original (Meyenn, 2005b, p. 1253): "der Laplace-Geist mißt so, daß für ihn zwei Pferde ebenso verschieden sind wie ein Pferd und ein Ochs".

What triggered this discussion about the Laplacian spirit is the notion of finite classes in biology, questioned by Pauli. Elsasser interestingly pointed out that Pauli interprets this notion in a non-empirical manner, while it intends to refer to nothing but a "quantitative formulation of empirical facts". Indeed, any class holds, within margins of error, members with common parameters. Measurements of these parameters for each member produces a different set of results. This condition holds for biology, while in physics "maximal measurements" reduce the samples of the set to pure states. Here Elsasser remarked that in biology the idea of a system being in a "pure state" makes no sense because (Meyenn, 2005b, p. 1265)

> "there exists no such thing as a one-to-one correspondence between the 'pure states' which make up the components of the ensemble . . . and the real, individual samples of physical systems which are represented by the ensemble. . . . the microscopic structure of a sample is not defined *except in terms of preceding interactions of the system with other systems*, and this leaves a great deal of indeterminacy."

Elsasser thus acknowledged biological systems as systems having a history in time. Because of this, it is "quite arbitrary where the observer stops his measurements".

In sum, with his more practically oriented remarks, Elsasser subscribed to the difference between biology and physics and seemed to support the value of the then upcoming approach taken in molecular biology. However, as soon as Elsasser himself adopted a more theoretical discourse, neo-Darwinism is once again questioned. For example, Elsasser (Meyenn, 2005b, p. 1267) concluded that "neo-Darwinism is incompatible with modern theoretical science, – *quite apart from any assumptions whatever about the nature of organisms.*" The incompatibility lies, on the one hand, in the second law of information theory. This laws states that in a mechanical-statistical system information cannot be generated *de novo*, while this is what neo-Darwinism says to happen. On the other hand, population genetics (used in evolutionary theory) is based on statistical analysis, which – according to Elsasser – is a matter of the second law of information theory.

While Elsasser admitted that he is not able to speak nor write about this issue to biologists themselves, Pauli is very supportive.[21] In a follow-up letter dated September 30, 1958, Pauli (see Meyenn, 2005b, p. 1272) speaks about

---

[21] In a long letter to Delbrück, dated October 6, 1958 (see Meyenn, 2005b, pp. 1279), Pauli recommends Elsasser's book and mentions how it reacts to evolutionary theory because "he also argues – very generally – in favor of the possibility of biological laws which cannot be reduced to physics" (German: "er vertritt – sehr allgemein – ebenfalls die Möglichkeit von biologischen Gesetzen, die sich nicht auf Physik zurückführen lassen"). He adds that "it is very risky if a physicist . . . alone writes a book about biology" (German: "es ist sehr gewagt, wenn ein Physiker . . . allein ein Buch über Biologie schreibt"). On April 12, 1958 (Meyenn, 2005b, p. 1143), Pauli also wrote to Weisskopf, quoting Delbrück who says, "the only avenue of progress in molecular biology today, as in atomic physics then, is to

"the good cause against neo-Darwinism". As an example, he refers to "the survival of the feathers": this[22]

> "reduces to the definition of 'feathers' as those which had the greatest likelihood to survive. One of them must be it, and this property can *generally* in no way be replaced by another one."

This last letter of Pauli to Elsasser also illustrates the importance of conceptual clarity. Pauli was confused by Elsasser's statement that in a mechanical-statistical system information cannot be generated *de novo*. Pauli (see Meyenn, 2005b, pp. 1271–1272) thought this principle was false[23] if not identical to the second law of thermodynamics, because negative entropy is proportional to information. Hence, it becomes incomprehensible why an organism cannot generate new information, while actually the organism looses information through metabolic reactions. Once again, this example shows how Pauli uses a mathematical conception to think about biology.

In Elsasser's response (see Meyenn, 2005b, pp. 1302–1303) (which Pauli never got to answer), it is explained how the information principle in question is not to be identified with the second law. Instead, Elsasser had a formal concept of heredity in mind: in case of any automaton A which generates an automaton B, (i) "B is not uniquely defined, but is any one of a set of possible automata that can be generated by A"; (ii) "the description of B is deducible from the description of A", while it is "not in general possible to recover a complete description of A (irreversible loss of information by noise)". Conversely, "a system which can generate 'information' not formally deducible from pre-existing information is not an automaton."[24] In a second letter, Elsasser (see Meyenn, 2005b, pp. 1311–1312) continues:

> "My term 'information' is merely a catchword for a set of interrelated parameters describing a structure, which description cannot be derived from,

---

develop better techniques and to do more ingenious experiments, to drive molecular biology along its traditional and possibly naive path. The difference between those who believe that this will be all plain sailing, and those who believe in the advent of an intellectual impasse to be resolved by a great revelation is like that between those who do not, and those who do, believe in a life hereafter" (Meyenn, 2005b, pp. 1144). On the basis of this quote, Pauli thinks "that in Delbrück some mystical background is 'constellated'" (German: "daß bei Delbrück irgendein mystischer Hintergrund 'konstelliert' ist"). It is difficult to interpret what Pauli had in mind here.

[22] German original (Meyenn, 2005b, p. 1272): "...reduziert sich auf die *Definition der 'feathers'*, diejenigen, welche die größte Chance hatten, zu überleben. Einer muß es ja sein, und diese Eigenschaft läßt sich *allgemein* in keiner Weise durch eine andere ersetzen."

[23] And he adds that "this was once said by a moslem about all other books with respect to the Koran" (German: "das hat schon einmal ein Moslem von allen übrigen Büchern in bezug auf den Koran gesagt"; see Meyenn (2005b, p. 1271).

[24] Elsasser argued, for simplicity's sake, that the information going into B from the environment is negligible, and if not, one may include it in A to begin with.

> say, the basic Hamiltonian (gross chemical composition) of the system considered. If this structure defines an automaton, then the automaton can only deteriorate, at best maintain or reproduce itself; it cannot create new automata *other* than those whose description is implicit in itself."

4 Conclusion: Transcendentalism *Versus* Naturalism, Irrationalism *Versus* Rationalism

The connecting threads underlying the diverse topics that Pauli discussed in his scientific correspondence are (i) the exploration of approaches alternative to the paradigm of rational and objective knowledge in science, and (ii) how quantum physics can offer assistance in this exploration.

On the basis of my reading of Pauli's speculations on science and biology, it appears that the "new reality" Pauli looked for remains to be interpreted in terms of a *natural and material reality* instead of a transcendent or supernatural reality. Pauli mainly intended to challenge an objective and observer-independent view on reality, in favor of a view in which the role of the observer forms part of the scientific endeavor. Acknowledging the enormous complexity of living organisms, understanding this role indeed becomes important because it is the observer who decides on what an organism is, how to cut it out from its environmental background, how to prepare it or aspects of it for observation with specific tools. This lesson today remains as important as ever.

What Pauli took less into account is that biological sciences also should be interpreted as a dynamic enterprise evolving in time. This allows us to see that in a continued and iterative process between experimental research and model building, the role of the observer becomes included in biological sciences. Still, where the traditional aim in science is to minimize or bring the influence of the observer under rational control, Pauli explored how the image of science would change when the role of the observer is fully included. The corresponding change of direction, Pauli concluded on the basis of his reflections, would be from rational to irrational. This reminds us of what Elsasser (1998, p. 129) argued:

> "One can think of reductionism as mainly an expression of what, in ordinary life, would be called a highly conservative viewpoint. It is based on a (no doubt largely unconscious) fear of the irrational. To appreciate this, we must recognize that such a fear of the essentially unknown, if not unknowable, differs from the fear of an anticipated event, say fear of the bite of a poisonous snake. It is ever so much easier to learn to control the fear of a specific event than it is to deal with an unknown and perhaps totally unknowable future."

The lesson Pauli learned was that rational scientific knowledge is limited and relative, not absolute. Like Elsasser argued, reality has aspects which are (as yet or in principle) unknown to us. Pauli did not simply accept this. Instead he tried to somehow grasp this irrationality and how it complements rationality. This gives away a lingering hope that a "new scheme" would provide a better, i.e. a more complete, picture of the reality humans experience. As such, Pauli's exploration remains bound to the traditional aim of science to reach the most adequate view on reality.

Throughout his exploration, Pauli did not eschew metaphysics, nor leaves to investigate what vitalistic or teleological approaches could have in store for our view on nature or reality. However, he realized (in a letter to Fierz of August 10, 1954) that "with these vague thoughts I reached the limits of what can be known today, and even came close to 'magic'.[25] With this, Pauli captured how his ideas are not just philosophical speculations but also unpronounced and vague. Consequently, he developed a philosophical modesty regarding the value of his ideas. This modesty is much less present in Elsasser's writings later on. For example, Elsasser (1998, p. 128) speaks of epistemology as "a form of reasoning which combines pure science with philosophy," and adds that (Elsasser, 1998, p. 128, italics added)

> "the ancient task of philosophy is to answer questions about the place of man in the universe; in doing so it uses reasoning as its primary tool but it is not confined to it. Philosophers have *successfully* used their *intuition* to fill out missing parts. If the universe cannot be fully comprehended by 'rational' means, then this use of intuition will not be preliminary but must be an inevitable part of the process of knowing."

The question which weight to give to philosophical intuition today stands central in a loaded debate[26] in the philosophical community. More specifically, the status of the reflections made in current philosophy of biology and the kind of relations that these entertain with the biological sciences is under discussion. Thus far, the debate seems to polarize between two positions: i.e. naturalistic philosophy (Callebaut, 2007) *versus* transcendental philosophy (cf. Kolen and Vijver, 2007). Whereas the former combines a growing acknowledgment of biological complexity in materialism and pragmatism, the latter stresses the (human) observer as key to any scientific understanding. This immediately reminds us of Pauli's search. However, that this transcendentalist stance today remains the marginal position makes us wonder in how far it has succeeded ever since in handling the problem Pauli faced: the problem of how to set intuition at work.

---

[25] German original (Meyenn, 1999, p. 745): "Ich bin mit diesen vagen Gedankengängen an die Grenze des heute Erkennbaren gekommen und habe mich sogar der 'Magie' genähert."

[26] For example, this was demonstrated at the Octavian discussion session on Transcendental versus Naturalistic Philosophy, which I organized at the 2007 ISHSSP-meeting in Exeter, UK, and in which little tolerance for transcendentalism could be detected.

My own contribution to this discussion aims to go *beyond* the dichotomy between naturalism and transcendentalism as presented by Callebaut and Kolen and Vijver (Speybroeck, 2007). The reason for this refers to Pauli's complementary view on rationality and irrationality. Both can indeed be said to play their part in any scientific endeavor. Hence, it comes down to see what their relation is about, instead of focusing on the one as if the other is non-existent, or to make a caricature of the other. Extending David Hull's view on the matter (Hull, 1969), I explore what it takes for philosophy of biology – as a discipline – to take both biology and philosophy seriously.

A first stress is put on philosophers not eschewing disciplinary self-reflection. This allows us to address in how far philosophical methodologies for analysis are *adequate* to tackle conceptual or other issues in biology. More generally, it allows us to address what can be expected from philosophical reflections. Do they hold an explanatory core ? Are they meant to challenge the intuitive ? And if so, on what grounds do they do so ? Clearing up such questions is a welcome exercise in the context of sustaining the often troubled (academic) relationship between philosophy and science.

A second and related stress is put on the development of a thorough acquaintance, not just with biological knowledge written down in academic papers, but also with the diverse epistemic cultures present in different biological disciplines. Such epistemic cultures have a specific historical background and combine aspects of both fundamental *and* funded science, experimental limitations *and* possibilities, metaphysics *and* pragmatism, ambitions *and* guiding academic structures. When taking these epistemic cultures into account, different lights may be thrown on where, when, and what kind of problems do or do not arise in biology, and how philosophically inspired reflections about these problems can provide input into the biological way of thinking.

I argue for an open and interested philosophy, away from ivory tower or armchair philosophy that runs by the conviction that major questions in biology (such as "what is life ?" and "can life be modeled ?") simply can be "thought". In my view, the transcendentalist philosophy defended by Kolen and Vijver (2007) runs more risk of being lured into such an armchair philosophy, giving the irrational free game under the defence that intuitions are as true (or even more true) as rational knowledge bound by strict experimental controls.

Pauli's correspondence may provide a modest, and therefore, welcome input to this debate in order to find a workable merger between naturalism and transcendentalism.

Acknowledgments

FWO-Flanders financially supported this research. I warmly thank Harald Atmanspacher for bringing Pauli's speculations about biology under my attention, and for offering me the opportunity to contribute to this volume. I am much indebted to Hans Primas for editorial suggestions and remarks.

References

Atmanspacher, H. and Primas, H. (2006): Pauli's ideas on mind and matter in the context of contemporary science. *Journal of Consciousness Studies* **13**(3), 5–50.

Callebaut, W. (2007): Transcendental niche construction. *Acta Biotheoretica* **55**, 73–90.

Crick, F. (1958): On protein synthesis. *Symposium of the Society for Experimental Biology* **12**, 138–163.

Crick, F. (1970): Central dogma of molecular biology. *Nature* **227**, 561–563.

Elsasser, W.M. (1958): *The Physical Foundation of Biology: An Analytical Study.* Pergamon, London.

Elsasser, W.M. (1998): *Reflections on a Theory of Organisms: Holism in Biology.* Hopkins University Press, Baltimore. (Originally published 1987, Editions Orbis, Frelighsburg, Quebec).

Enz, C.P. and Meyenn, K. von (eds.) (1994): *Wolfgang Pauli. Writings on Physics and Philosophy.* Springer, Berlin.

Espagnat, B. de (2006): *On Physics and Philosophy.* Princeton University Press, Princeton.

Fagot-Largeault, A., Rahman, S. and Torres, J.M. (eds.) (2007): *The Influence of Genetics on Contemporary Thinking.* Springer, Dordrecht.

Gershon, D. (2002): Silicon dreams in the biology lab. *Naturejobs* **417**, 4–5.

Helrich, C.S. (2007): Is there a basis for teleology in physics? *Zygon* **42**, 97–110.

Holliday, R. (1996): DNA methylation in eukaryotes: 20 years on. In: Russo, V.E.A., Martienssen, R.A., and Riggs, A.D. (eds.), *Epigenetic Mechanisms of Gene Regulation.* Cold Spring Harbor Laboratory Press, New York, pp. 5–27.

Hull, D.L. (1969): What philosophy of biology is not. *Journal of History of Biology* **2**, 241–268.

Ideker, T., Galitski, T. and Hood, L. (2001): A new approach to decoding life: systems biology. *Annual Review of Genomics and Human Genetics* **2**, 343–372.

Jablonka, E. and Lamb, M. (1995): *Epigenetic Inheritance and Evolution: The Lamarckian Dimension.* Oxford University Press, Oxford.

Keller, E.F. (1995): *Refiguring Life: Metaphors of Twentieth-Century Biology.* Columbia University Press, New York.

Kitano, H. (2001): Systems biology: Towards system-level understanding of biological systems. In: Kitano, H. (ed.), *Foundations of Systems Biology.* MIT Press, Cambridge, MA, pp. 1–38.

Kolen, F. and Vijver. G. van de (2007): Philosophy of biology: naturalistic or transcendental? *Acta Biotheoretica* **55**, 35–46.

Laurikainen, K.V. (1988): *Beyond the Atom. The Philosophical Thought of Wolfgang Pauli.* Springer, Berlin.

Leene, J. van, Stals, H., Eeckhout, D., Persiau, G., Slijke, E. van de, Isterdael, G. van, Clercq, A. de, Bonnet, E., Laukens, K., Remmerie, N., Hendrickx, K., Vijlder, T. de, Adbelkrim, A., Pharazyn, A., Onckelen, H. van, Inzé, D., Witters, E. and Jaeger, G.de (2007): A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis thaliana. Molecular and Cellular Proteomics* **6**, 1226–1238.

Lewin, B. (1998): The mystique of epigenetics. *Cell* **93**, 301–303.

Lewontin, R.C. (2001): Gene, organism and environment: A new introduction. In: Oyama, S., Griffiths, P.E., and Gray, R.D. (eds.), *Cycles of Contingency. Developmental Systems and Evolution.* MIT Press, Cambridge, MA, pp. 55–58.

Meyenn, K. von (ed.) (1993): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band III: 1940–1949.* Springer, Berlin.

Meyenn, K. von (ed.) (1999): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil II: 1953–1954.* Springer, Berlin.

Meyenn, K. von (ed.) (2005a): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil IV-A: 1957.* Springer, Berlin.

Meyenn, K. von (ed.) (2005b): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil IV-B: 1958.* Springer, Berlin.

Mitchell, S.D. (2003): *Biological Complexity and Integrative Pluralism.* Cambridge University Press, Cambridge.

Oyama, S., Griffiths, P.E. and Gray, R.D. (eds.) (2001): *Cycles of Contingency. Developmental Systems and Evolution.* MIT Press, Cambridge, MA.

Pauli, W. (1948): Editorial. *Dialectica* **2**, 307–311.

Pauli, W. (1954): Naturwissenschaftliche und erkenntnistheoretische Aspekte der Ideen vom Unbewussten. *Dialectica* **8**, 283–301.

Pauli, W. (1957): Phänomen und physikalische Realität. *Dialectica* **11**, 36–48.

Sarkar, S. (1996): Biological information: A skeptical look at some central dogmas of molecular biology. In: Sarkar, S. (ed.), *The Philosophy and History of Molecular Biology: New Perspectives.* Kluwer, Dordrecht, pp. 187–231.

Schäfer, L. (2006): Quantum reality and the consciousness of the universe. *Zygon* **41**, 505–532.

Smith (2005): The "hard problem" and the quantum physicists. Part 1: The first generation. *Brain and Cognition* **61**, 181–188.

Speybroeck, L. van (2000): The organism: a crucial genomic context in molecular epigenetics? *Theory in Biosciences* **119**, 187–208.

Speybroeck, L. van (2002): From epigenesis to epigenetics: The case of C.H. Waddington. In: Speybroeck, L. van, Vijver, G. van de, and Waele, D.D. (eds.), *From Epigenesis to Epigenetics: The Genome in Context.* Annals of the New York Academy of Sciences, New York, pp. 61–82.

Speybroeck, L. van (2007): Philosophy of biology: About the fossilization of disciplines and other embryonic thoughts. *Acta Biotheoretica* **55**, 47-71.

Speybroeck, L. van, Vijver, G. van de and Waele, D.D. (2007): "Epi-geneticization": Where biological and philosophical thinking meet. In: Fagot-Largeault, A., Rahman, S., and Torres, J.M. (eds.), *The Influence of Genetics on Contemporary Thinking.* Springer, Dordrecht, pp. 115–133.

Torretti, R. (1999): *The Philosophy of Physics.* Cambridge University Press, Cambridge.

Vidal, M. (2001): A biological atlas of functional maps. *Cell* **104**, 333–339.

Vijver, G. van de and Speybroeck, L. van (2006): Philosophy of science meets biological complexity. Review of "Biological Complexity and Integrative Pluralism". *Biological Theory. Integrating Development, Evolution, and Cognition* **1**, 104–106.

Vijver, G. van de, Speybroeck, L. van and Vandevyvere, W. (2003): Reflections on complexity in biological systems: Kant and beyond. *Acta Biotheoretica* **51**, 101–140.

Vijver, G. van de, Speybroeck, L. van and Waele, L. de (2002): Epigenetics: A challenge for genetics, evolution and development? In: Speybroeck, L. van, Vijver, G. van de, and Waele, D.D. (eds.), *From Epigenesis to Epigenetics: The Genome in Context*. Annals of the New York Academy of Sciences, New York, pp. 1–6.

Index