

Abduction, Competing Models and the Virtues of Hypotheses

H. G. Callaway

Abstract This paper focuses on abduction as explicit or readily formulatable inference to possible explanatory hypotheses—as contrasted with inference to conceptual innovations or abductive logic as a cycle of hypothesis, deduction of consequences and inductive testing. Inference to an explanation is often a matter of projection or extrapolation of elements of accepted theory for the solution of outstanding problems in particular domains of inquiry. I say, “projection or extrapolation” of accepted theory, but I mean to point to something broader and suggest how elements of accepted theory constrain emergent models and plausible inferences to explanations—in a quasi-rationalistic fashion. I draw illustrations from quantum gravity below just because there is so little direct evidence available in the field. It is in such cases that Peirce’s discussions of abduction provide the most plausible support for the idea of a logic of abduction—as inference to readily formulatable explanatory hypotheses. The possible need for conceptual innovation points to limits on the possibility of a logic of abduction of a *more* rationalistic character—selecting uniquely superior explanations. Abduction conceived as a repeating cycle of inquiry also points to limits on our expectations for an abductive logic. My chief point is that the character of inference to an explanation, viewed below as embedded within arguments from analogy, is so little compelling, as a matter of logical form alone, that there will always be a pluralism of plausible alternatives among untested hypotheses and inferences to them—calling for some comparative evaluation. This point will take us to some consideration of the virtues of hypotheses—as a description of the range of this pluralism.

H. G. Callaway (✉)
Philadelphia, PA, USA
e-mail: HGCallaway@live.com

1 Conceptual Constraint on Hypothesis

The constraint of accepted theory and pre-existing conceptual resources upon inference to explanatory hypothesis cannot be absolute. In proposing a new explanatory hypothesis, one typically seeks to meet problems in accepted theory, account for some anomaly, by way of supplementing and/or modifying theory or elements of belief, and these proposals may involve innovations in the conceptual resources of the theory. Yet, this is far from saying that the pre-existing theory in a domain and the conceptual resources employed in it will have no influence on plausible inferences to explanatory hypotheses.

Consider the abstract, bare-bones scheme which Peirce provides for abductive inference in his late writings¹:

The surprising fact, C, is observed;
But if A were true, C would be a matter of course.
Hence, there is reason to suspect that A is true.

I will pass for the moment on the question of whether such inference requires a distinctive logic, properly so called, for its elucidation, though I acknowledge that this kind of inference welcomes elucidation and that similar inferences form a class concerning which study is advised. Whether or not there is a logic of abduction,² I think there is an *art* of inference to explanatory hypotheses and that this is worthy of study and attention. Still the plausibility of acknowledging an art of inference to hypotheses is part of the plausibility of denying a distinctive logic of abduction. As formulated here, abduction is a matter of cognitive expectations. First of all, some fact C is said to be “surprising,” which is to say that C in some manner fails to accord with established expectations. In the second premise, the supposition is that if the hypothesis A were true, then “C would be a matter of course,” which is to say that C would no longer be surprising, but instead would accord with expectations arising in consequence of the supposed truth of hypothesis A. What lends support to hypothesis A in the conclusion is that the surprising character of C would be removed and C would become consistent with envisaged expectations arising on the supposition of the hypothesis. To evaluate this form of inference we have to know whether and how the truth of the premises would render the conclusion plausible. What, then, changes the relevant expectations?

The first requirement is to understand what it is that creates the expectations, including the initially surprising character of fact C and the removal of the surprising

¹ See Peirce, C.S. (1903) [17] “Pragmatism as the Logic of Abduction,” in *The Essential Peirce, Selected Philosophical Writings*, Vol. 2 (1893–1913), p. 231.

² Regarding alternative approaches to the topic of the logic of abduction, see Aliseda, A. (2006) [1] *Abductive Reasoning. Logical Investigations into Discovery and Explanation*; Gabbay, D.M. and J. Woods (2005) [7] *The Reach of Abduction, A Practical Logic of Cognitive Systems*, 2 vols.; and Magnani, L. (2009) [14] *Abductive Cognition, The Eco-Cognitive Dimensions of Hypothetical Reasoning*.

character on the supposition of hypothesis A. The answer which seems most reasonable is that in the two cases, we envisage or presuppose some accepted, or prospective, context of theoretical understanding (or belief) relevant to the fact C, of which C is at first not, and afterward becomes, an expected part. We might imagine, for instance, that fact C presents a counter-example to an accepted theory T_1 over domain D, or that it is not encompassed by theory T_1 ; and in consequence, fact C is surprising to those whose expectations regarding domain D are structured by their acceptance of theory T_1 . Correspondingly, to say that if hypothesis A were accepted, then “C would be a matter of course,” suggests that there is some alternative formulatable theory T_2 , including A, over domain D, such that in the simplest case, C or a related conditional with C as consequent, is a logical implication of theory T_2 . In short, the idea is that it is always a theoretical context, or (weakly or firmly held) beliefs (which can be idealized as a theory), including some typical patterns of inference among its concepts or terms, which structures and is suited to create or remove the conceptually relevant expectations involved in the “surprising” or “matter of course” character of particular observed facts of experience.

This is to say that given particular intellectual configurations involving accepted belief or theory and problematic evidence of facts relevant to, or included within, the same domain as the accepted theory—such configurations—will always create some needed conceptual constraint upon possible hypotheses designed or functioning to deal with the problematic aspects of the theoretical situation. For, it is the concepts and patterns of inference of some particular theory, or idealization of existing belief, which renders some fact C “surprising” in the first place, and plausible constraint upon alternative hypotheses and corresponding alternative belief or theory must seek to preserve relevant patterns of expectation while removing the particular surprise. Preserving relevant patterns of expectation, so far as possible, means preserving the explanatory accomplishment of the theory or belief system theretofore accepted. Here we need to notice, too, that consistent with the argument for hypothesis A given above, we may reasonably suppose that any of several alternative arguments are open to us:

The surprising fact, C, is observed;
 But if A' were true, C would be a matter of course.
 Hence, there is reason to suspect that A' is true.

So long as we deal with these matters in this abstract fashion, we can consistently suppose that we have any number of similar possible inferences, each putting forth some particular possible hypothesis, A', A'', A''', etc. The pattern or generalized form of inference Peirce ascribes to abduction is not plausibly regarded as a matter of inference to some unique “best explanation,” and is instead better understood as a pattern of inference to any of several possible explanations. This in turn suggests that it is not the *pattern of inference* which “wears the pants,” in actual and successful inferences of this sort, and that we have to access the actual content and concrete problems of particular fields of inquiry in order to make much sense out of inference to explanatory hypotheses. In short, we need to examine, and depend

upon, the actual patterns of conceptual expectations arising in particular contexts of inquiry: it is these concrete expectations which “wear the pants” in any actual inference to an explanation.

It is reasonable to expect such a result, and perhaps the point will not be very widely questioned. Explanations are always embedded in particular conceptual frameworks or theoretical systems. If a virus V is the cause of disease D, say, the common cold, and we can, in fact, re-describe virus V, as the last thing mentioned by Jones, then it follows that the last thing mentioned by Jones is the cause of disease D. More generally, if A is the cause of B, then A remains the cause of B under any alternative description which picks A out. However, an explanation depends on particular conceptual resources and theoretical context, an explanation makes use of preferred or salient forms of description linked by theoretical context to particular consequences; and alternative characterizations of the things we refer to in explanations do not carry the same explanatory force or weight. We might explain someone catching cold by contact with virus V, but it won't do to offer explanation in terms of contact with the last thing mentioned by Jones. Our accepted theoretical concepts have a certain inherited salience or grounded projectability, due to their roles in successful explanation and their comprehension of a range of evidence. We naturally want to preserve them so far as possible in the problematic situation. Regarding things mentioned by Jones, in general, we have no relevant generalizations.

There will always be a range of explanatory innovations that may be proposed, regarding unsolved problems, running from the more conservative to the less conservative; and it is important, in ruling out “wild guessing,” that attention be initially directed to more conservative proposals. To say that there is at least a quasi-logic, or art of abduction, is to resist over-emphasis on the idea that “hypothesis is guesswork” or that it is merely guesswork. If there is an important distinction between “guesswork” and “enlightened guess work,” as Quine and Ullian maintain, in *The Web of Belief*, then this points to the possibility of systematic and comparative evaluation of the better and worse among unverified hypotheses³—thought alternative abductive arguments share the same abstract form.

2 Quantum Gravity, Analogy and Hypothesis

Contemporary inquiries and discussions of the relationship between general relativity and quantum mechanics are of particular interest in the present context, just because empirical evidence has been so scarce. The gravitational interaction, though cumulative over long distances, is much weaker than any of the other interactions, and because of this it is exceedingly difficult to institute experiments

³ “In a word, hypothesis is guesswork,” say Quine and Ullian (1978) [24] in *The Web of Belief*, second ed., p. 65, “but it can be enlightened guesswork.”

sufficiently delicate to measure effects requiring the precision of a quantum theory of gravitation.⁴ The conflicts between general relativity and quantum mechanics are of a theoretical rather than a more empirical nature: basic in this is the conflict between the smooth continuities of the gravitational field of relativistic space-time in contrast to the bubbling discreteness of quantum fields. As Brian Greene has put the point, “The notion of a smooth spatial geometry, the central principle of general relativity, is destroyed by the violent fluctuations of the quantum world on short distance scales.”⁵

In *Objective Knowledge* (1979), Karl Popper provided a point of reference in opposition to methodological conservatism when he characterizes his conception of conjecture in relation to “the method of science”: “*The method of science*,” he says, “*is the method of bold conjectures and ingenious and severe attempts to refute them*;”⁶ This implies that in the comparative evaluation of a pair of new hypotheses, as possible modifications of some pre-existing theory, we should generally prefer that new theory, containing an hypothesis, which has greater logical comprehension, implying a larger range of testable consequences. Einstein’s bold innovations regarding space-time and matter in motion are a plausible model here.

While Stephen Hawking expresses sympathy for Popper’s falsificationism, he says, too, that “In practice, it seems that one develops a new theory which in truth is only an extension of the old.”⁷ Brian Greene writes that “rather than trying through one leap, to incorporate all we know about the physical universe in developing a new theory, it is often far more profitable to take many small steps that sequentially include the newest discoveries from the forefront of research.”⁸ This claim compares more favorably with Popper’s emphasis on “trial and error” and the “piece-meal approach,” in *The Poverty of Historicism* (1957) and equally with Popper’s emphasis on simplicity in *The Open Universe* (1982).⁹ “The method of science,” he says there, “depends upon our attempts to describe the world in simple theories.” In contemporary physics, the beauty and generality of Einstein’s physics of space-time and gravitation seems to be at war with the precision and predictive success of quantum mechanics; and Einstein’s later dream of a unified field theory, though arguably more plausible during Einstein’s years at Princeton—as a conservative modification of his physics of gravitation, space-time and matter aiming to integrate electromagnetism—now counts as a proposal far wide of the mark, since the quantum-mechanical approach has subsequently encompassed three of the four fundamental forces from the opposite direction—as a quantum field theory. What

⁴ See, e.g., Feynman, Richard (1985) [6] *QED: The Strange Theory of Light and Matter*, p. 151.

⁵ Greene, Brian (1999) [9] *The Elegant Universe*, p. 129.

⁶ Popper, Karl (1979) [23] *Objective Knowledge*, revised edition, p. 81.

⁷ See Hawking, Stephen (2006) [11] *A Briefer History of Time*, p. 20 in the German edition.

⁸ Greene (1999) [9] *The Elegant Universe*, p. 121.

⁹ See Popper, Karl (1957) [21] *The Poverty of Historicism*, p. 75; Popper, Karl (1982) [22] *The Open Universe*, p. 44.

counts as a more reasonable hypothesis or general direction of inquiry, changes with, and thus depends upon, the specific details of our context of knowledge.¹⁰

In a somewhat similar way, recent critics of super-symmetry, string theory and their developments, theories aiming for a unified quantum mechanical approach to the four known forces including gravitation, have become increasingly strident in recent years, pointing to a persistent vagueness which has yielded no significant predictions over a period of some 20 years.¹¹ The contemporary conflicts and alternatives chiefly turn on varieties of string theory, based in particle physics, and versions of quantum gravity which require no background metric and are more closely related to relativity. Lacking a logic of abduction, we, and the physicists, do best to listen to many voices.

The unification of three of the four known forces as a quantum field theory now promises further developments from that direction. The genius of Einstein remains beyond doubt, of course; and in fact, Einstein foresaw the general conflict. As early as 1916 he wrote that “because of the intra-atomic movement of electrons, the atom must radiate not only electromagnetic but also gravitational energy, if only in minute amounts,” thus “...it appears that the quantum theory must modify not only Maxwell’s electrodynamics but also the new theory of gravitation.”¹² There has been some recognition of the tensions between quantum mechanics and Einstein’s physics of space, time and gravitation from the very start.

In the history of theory in quantum gravity, in view of the lack of access to the relevant energies and related empirical evidence, much of the development has been by way of analogies with developments in related fields of study. Quantum mechanical considerations, it was thought, would have to modify Einstein’s theory of the gravitational field, and the theory of the gravitational field would have its impact upon quantum mechanics. But how exactly? Summarizing some of the history, Dean Rickles, in a recent textbook on the philosophy of physics, sees the development as a matter of arguments by analogy:

When quantum gravity was finally studied in a systematic way, it was undertaken to a considerable degree on the basis of analogies with other fields. Inferences were made (not always soundly) on the basis of these analogies to physics at the Planck scale. Later work revealed the inadequacies in this analogical reasoning.¹³

It is sometimes held that arguments from analogy are arguments from the character of one particular to that of another similar particular, and that would seem to

¹⁰ See the opening discussion of the theoretical conflict in the middle of the last century in Stachel, John (1999) [29] “Comments” in Cao, Tian Yu, ed. (1999) [4], pp. 233–234.

¹¹ The related criticisms are forcefully summarized in Woit, Peter (2007) [31] *Not Even Wrong, The Failure of String Theory and the Continuing Challenge to Unify the Laws of Physics*. See also Smolin, Lee (2006) [27] *The Trouble with Physics*.

¹² Einstein, Albert (1916) [5] “Approximate integration of the field equations of gravitation,” in Engel, A. ed. (1997) *The Collected Papers of Albert Einstein*, Vol. 6., p. 209. Quoted in Rickles, Dean (2008) [25] *Companion to Contemporary Philosophy of Physics*, p. 285, n. 39.

¹³ Rickles, Dean (2008) [25] *Companion to Contemporary Philosophy of Physics*, p. 284.

introduce a significant basis for contrast between arguments from analogy and the presuppositions of abductive inferences.¹⁴ But on the other hand, the analogies involved in quantum gravity appear to be analogies between better established theories and more speculative theories—or conjectural models possibly contributing to such speculative theories. So, the idea might be, for instance, that the gravitational force must be like the electromagnetic force, both involve fields and attractions of one thing to another, and in consequence, since we recognize the photon, with its wave/particle duality, as carrying the electromagnetic force, it seems reasonable to suppose that there must be a similar entity, the graviton, with its own wave/particle duality, to carry the gravitational force.

That this is part of a significant argument from analogy in the history of quantum gravity is evident from that fact that it has been responded to by use of arguments based on important disanalogies, e.g., the argument put forward by the Soviet physicist M. P. Bronstein in the 1930s and later brought to the attention of the wider world.¹⁵ The chief idea arising from related developments is that quantum gravity cannot be formulated by direct analogy with quantum electrodynamics but instead that the unique features of gravitation require special treatment in which some generalization of quantum field theory is needed—one which is applicable in the absence of a fixed background metric. This would be a version of quantum field theory based on a different analogy—an analogy to the dependence of the space-time metric on mass-energy content as in Einstein's physics.

I am sure that many similar illustrations could be provided of projections or extrapolations regarding quantum gravity, and this thought is encouraged by viewing abductive inference as guided by existing theoretical/conceptual regularities. This is to say, in effect, that knowledge of the relevancy, or weight, to be assigned to particular similarities is not known *a priori*, but arises instead from the actual establishment of theory and related habits or patterns of inference—based in the specifics of the theories in question. Still, advances in knowledge can establish new relevancies and disrupt the old, and in consequence, though we are never without some sense for the difference between strong and weak analogies, this is bound to evolve in degree with the growth of knowledge. Such context dependency of the plausibility of analogical and abductive reasoning helps us understand a famous quotation illustrating misplaced early confidence from the likes of Werner Heisenberg and Wolfgang Pauli—in their first 1929 paper on quantum electrodynamics:

¹⁴ Contrast the discussion in Thagard, Paul and C. Shelly (2001) [30] “Emotional Analogies and Analogical Inference,” in Gentner, Dedre et al. (eds.) (2001) [8] *The Analogical Mind*, pp. 336–337.

¹⁵ See the discussion of Bronstein's work in Stachel, John (1999) [28] “Introduction: Quantum field theory and space-time,” in Cao, Tian Yu, ed. (1999) [4] *Conceptual Foundations of Quantum Field Theory*, pp. 171–173.

Quantization of the gravitational field, which appears to be necessary for physical reasons, may be carried out without any new difficulties by means of a formalism fully analogous to that applied here.¹⁶

Obviously, the analogy was not without its appeal, given what they knew or thought most significant at the time; but the present context of physical knowledge renders it much less plausible. Standard quantum mechanics presupposes a fixed background metric—a scheme of locations and times at which particles interact—, while the metric is internal to the gravitational field.

Although it may do some damage to the Peircean scheme of philosophical triads,¹⁷ to see abductive inference as an abstract element of arguments from analogy, this is suggested by emphasis on the role of theoretical context in inferences to explanatory hypotheses. On somewhat similar grounds, comparative evaluation of competing (untested) hypotheses, might be viewed as a matter of stronger and weaker analogies—i.e., comparative evaluation of alternative models which arise from reinterpretation of various living theories. In particular cases, or perspectives of discussion, a moment of abductive inference may stand out, while in other cases some embedding analogy stands in the forefront of attention.

3 Penrose on Quantum Gravity

On of the most fascinating recent proposals regarding quantum gravity is the idea from Roger Penrose that quantum state reduction is a *gravitational* phenomenon. “I belong to the general school of thought that maintains,” he writes, “that the phenomenon of quantum state reduction is a *gravitational* phenomenon,” and moreover, as he further puts his claims, “essential changes are needed in the framework of quantum mechanics in order that its principles can be adequately married with the principles of Einstein’s general relativity.”¹⁸

One way to understand this proposal is to ask why we do not see macroscopic superpositions. If quantum mechanics is a general description of physical reality, which we suppose it is, and subatomic particles and related things (at least up to the size of 60-atom, carbon “Buckyballs”) can be placed into quantum mechanical superpositions (the state of being in more than one place at the same time) then this creates the expectation (often dampened in various interpretations of the

¹⁶ Heisenberg, Werner and Wolfgang Pauli (1929) [13] “*Zur Quantenelektrodynamik der Wellenfelder*,” *Zeitschrift für Physik*, 56, pp. 1–61. Quoted in Stachel (1999) [28] “Introduction,” p. 168.

¹⁷ See my discussion of Peirce’s triadic category theory in Callaway, H.G. (2008) [3] “A Role for Peirce’s Categories?” in Callaway, H.G. (2008) *Meaning without Analyticity, Essays on Logic, Language and Meaning*, pp. 183–192.

¹⁸ See, e.g., Penrose, Roger and P. Mercer (1998) [20] “Quantum Computation, Entanglement and State Reduction,” *Philosophical Transactions, Mathematical, Physical and Engineering Sciences*, Vol. 356, No. 1743, p. 1932.

formalism) that the objects of every-day life should be capable of entering into superpositions—though this is not observed.

One way to envisage the phenomenon of superposition is to consider the classical double-slit experiments. We set up a light source, a laser for example, and shine it in the direction of a screen which will brighten when the light encounters it, and between the screen and the source, an opaque barrier is placed which has two parallel slits through which the light may pass. If either slit is used alone, by covering the other, then the result is a lighted column behind that slit. However, if both slits are opened, then the result is a broad interference pattern of darker and brighter bars across the screen. When only one slit is available, the light seems to travel as discrete particles, coming out more or less directly behind the open slit. However, when both slits are available, then the light behaves in a wave-like fashion. What comes through each slit interferes with what comes through the other—producing a characteristic pattern of dark and light bands. The results sometimes suggest particles, sometimes waves. These results persist, moreover, if electrons are used instead of light, or even if molecules are used; and they persists if the photons or electrons or molecules are sent out one by one. It is as though each particle is in two places at once (in superposition), goes through both slits and then interferes with itself. If the two slits are open, though the particles are sent through one at a time, then the interference pattern slowly accumulates on the screen, while if only one slit is open, then all seems to be a matter of particles going through the one open slit. Even more curious, if a detector is installed to determine which of the two slits a particle traverses, then the interference pattern disappears: collapse of the wave function.¹⁹

On the assumption that quantum state reduction, the collapse of the wave function, as in measurement, is a *gravitational* phenomenon, the absence of observed macroscopic objects in superposition would seem to make good sense. The greater the mass of an object placed in superposition, the *less likely* it is that such a superposition will be stable for any length of time—superpositions of increasing mass should be increasingly rare. We have an inference to a hypothesis.

Penrose aims to avoid the Copenhagen and “many worlds” interpretations and offer a realist conception. State reduction, or the collapse of the quantum mechanical wavefunction, is a result of gravitational interactions in nature, requiring no observer—it is a result of physical interaction. On his view, gravity pulls objects back into a single location without need of an observer (or multiple worlds), and the more massive an object in superposition, then the shorter the time of any stable superposition.

Penrose is also much concerned to emphasize a methodological conservatism: “the complete agreement that the standard quantum formalism has with all experiments to date;” it is only in newly proposed experimental situations that Penrose’s proposal could be verified or experimentally disconfirmed. He also

¹⁹ Compare the discussion in Penrose (2004) [19] *The Road to Reality, A Complete Guide to the Laws of the Universe*, pp. 504–505.

builds on “a certain already existing conflict between the fundamental principles of general relativity and of quantum mechanics.”²⁰ The conflict is implicit in the “measurement problem” of quantum mechanics “which is to comprehend how, upon measurement of a quantum system, this (seemingly) discontinuous “R-process” [reduction of the state vector or collapse of the wavefunction] can come about”—given the expectations connected with the “U-process,” the uniform linear evolution of an encompassing quantum system “solely according to the Schrödinger equation.”²¹

In order to preserve the expected uniform evolution called for by the Schrödinger equation, some quantum physicists have gone so far as to suppose that multiple, unobservable worlds are required, parallel to ours without subsequent observable interaction—in which, collectively, everything that can happen in accordance with the equation does happen. The Copenhagen interpretation, in contrast, supposes that it is observation which brings about state reduction and a measurable result or outcome.

In contrast to earlier interpretations or approaches to quantum mechanics, Penrose argues for reduction as an objective phenomenon, the wavefunction is a physical wave, and that “present-day quantum mechanics is a limiting case of some more unified scheme, whereby the U and R procedures are *both* to be approximations to some new theory of physical reality.”²² Though he does not propose anything like a full theory of state reduction, he offers a model which is intended to constrain any broader theory. This builds from a “basic conflict” which Penrose sees between “Einstein’s covariance principle and the basic principles of quantum theory, as they related to stationary states of superposed gravitational fields.”²³

Einstein’s covariance principle tells us that the forms of physical laws are invariant under arbitrary transformations of coordinate systems. There is no privileged coordinate system. On the other hand, Penrose asks us to imagine a situation in which a very small lump of some rigid material has been placed in a quantum superposition of two positions —this thought experiment is to be regarded as a “an inanimate version of ‘Schrödinger’s cat’.”²⁴ The argument is that, ignoring gravitation, “the two alternative locations of the lump will each be stationary states,” and in consequence of quantum mechanical considerations arising from the Schrödinger equation, the linear combination of the two is also a stationary state. The overall configuration does not evolve in time.

²⁰ Penrose, Roger and P. Mercer (1998) [20] “Quantum Computation, Entanglement and State Reduction,” pp. 1932–1933.

²¹ Penrose, Roger (1996) [18] “On Gravity’s Role in Quantum State Reduction,” *General Relativity and Gravitation*, Vol. 28, No. 5, pp. 581–582.

²² *Ibid.*, p. 583.

²³ *Ibid.*, p. 584.

²⁴ *Ibid.*, pp. 584–585.

Next we are to consider the gravitational fields associated with the two superposed positions of the rigid lump of material. There is also a superposition of the associated gravitational field, and the reader begins to see how Penrose intends to argue for a “gravitational role in state-vector reduction.”²⁵ “But the principle of general covariance denies any significance to particular coordinate systems,” he argues, and hence “it asserts that there should be no preferred pointwise identification between two different spacetimes.”²⁶ This is a problem, because quantum field theory, in its usual forms, assumes that a background metric is given, and this assumption is needed to make calculations about possible outcomes. Something else is required.

Given a role for gravity in reduction of the superposition, then this must be in one or the other direction, though neither is privileged. That is the conflict or tension which Penrose envisages. It is partly a conflict between the expectations of a stationary state created by quantum mechanics, in light of the prospect of gravitational interactions, and partly a matter of the fact that there is uncertainty of location related to the non-privileged alternative reductions. It takes energy to sustain the superposed fields (which are slight distortions of space-time generated by the superposition of the associated mass), and the larger the masses involved the more energy it takes. Penrose’s proposed solution is to say that the supposedly stationary state of superposition is in reality somewhat like an unstable atom or particle which has a probability of decay to be calculated by consideration of its mass-energy uncertainty. As he elsewhere puts the point, there is a “fundamental energy uncertainty, EG” of the superposition, and,

The next step is to invoke a form of Heisenberg’s uncertainty principle (the time/energy uncertainty relation)... . It is a familiar fact, in the study of unstable particles or unstable nuclei (such as Uranium U_{238}) that the average *lifetime* T , having an inbuilt time uncertainty, is reciprocally related to an energy uncertainty, given by $\hbar/2T$. Now we are going to think of our superposed state $|\Psi\rangle = w|\chi\rangle + z|\phi\rangle$ as being analogous to this, itself being unstable, with a lifetime T_G that is related, by Heisenberg’s formula, to the fundamental energy uncertainty E_G [of the superposition]. According to this picture, any superposition like $|\Psi\rangle$ would therefore decay into one or the other constituent states, $w|\chi\rangle$ or $z|\phi\rangle$, in an average time scale of $T_G \approx \hbar/E_G$.²⁷

The superposition, $|\Psi\rangle$, though stable, a standing wave, in accordance with the Schrödinger equation, is said to be *analogous* to an unstable nucleus or subatomic particle. The average lifetime, or rate of decay, is then proportional to (a function) \hbar , of the Planck constant, divided by the energy uncertainty of the superposition of the

²⁵ *Ibid.*, p. 583.

²⁶ Penrose, Roger (2004) [19] *The Road to Reality*, p. 850.

²⁷ *Ibid.*, p. 853; cf. Rovelli, Carlo (1999) [26] “‘Localization’ in Quantum Field Theory,” in Cao, Tian Yu, ed. (1999) [4], p. 221. Calling for a “theory of quantum geometry,” Rovelli argues that given “the need to accommodate the superposition principle, and thus the possibility of quantum superposition of distinct geometries,” there can be no “well defined metric structure” for “all quantum states of the theory.” This is to reject simple presupposition of a fixed background metric in a fully adequate quantum field theory.

gravitational field—which is a function of the mass of the object placed in superposition.

Since both \hbar , a function of the Plank constant, and E_G , the energy uncertainty of the superposition, are very small, dividing through these two very small quantities, the average lifetime of the superposition, T , is expected to be measurable under plausible experimental conditions. In this way, Penrose’s proposal escapes the constraint of extremely high energies otherwise taken to be required to probe the effects of quantum gravity.

In a somewhat earlier formulation, Penrose writes,

To compute the decay time, according to this proposed scheme, consider the energy E that it would cost to pull away one instance of the mass, moving it out away from coincidence, in the gravitational field of the other, until the two mass locations provide the mass superposition under consideration. I propose that the time scale of the collapse of the state vector of this superposition is of the order of

$$T \sim \hbar/E$$

For a nucleon, this would be nearly 10^8 years, so the instability would not be seen in existing experiments. However, for a speck of water of 10^{-5} cm in size, the collapse would take about 2 hours. If the speck were 10^{-4} cm, the collapse would take about 1/10 sec, whereas for 10^{-3} cm size, the collapse of the state vector would take place in only some 10^{-6} sec.²⁸

Again, a superposition may be compared to an unstable nucleus, and the greater the mass of the object in superposition, the greater its instability and the shorter its average lifetime T —as with a transuranic nucleus of great atomic weight. It is clearer in this passage, why the phenomenon in question would not yet have been observed.

Though we have been viewing what is going on in Penrose’s proposal as an inference to a hypothesis, in extended stretches of this story, namely the hypothesis that gravitation has a role in state reduction, it appears in the wider context that this hypothesis arises from a certain analogy. This is an analogy between well established quantum mechanical theories of the decay of unstable nuclei and state reduction—functioning as a constraint on the unification of quantum mechanics with general relativity. Part of the appeal of this reasoning arises, because the promise of constraint on possible unification is rare. Though we are dealing with an as yet untested model,²⁹ intended to constrain some envisaged, but not yet formulated theory, the reasoning clearly enters into a small charmed circle of similar efforts, including, Stephen Hawking on black holes and “Hawking radiation”—significantly bridging the gap between quantum mechanics and general relativity. The supposition of Hawking radiation also arises as a gravitational

²⁸ See Penrose in Hawking, Stephen and Roger Penrose (1996) [12] *The Nature of Space and Time*, p. 72.

²⁹ See the plan for an experimental design to test Penrose’s hypothesis in Marshall, W.C., Simon, R., Penrose, R., et al. (2002) [15] “Towards quantum superposition of a mirror,” *Quantphl* 0210001v1, pp. 1–5.

effect, since pairs of virtual particles are separated—at the event horizon of black holes. It is in this way we can understand how black holes radiate.³⁰

That we find an analogy at the heart of Penrose’s proposal suggest a significant dependence of inference to a hypothesis upon analogical reasoning—as an intermediate context supportive of such inference. This creates the expectation that particular abductive inferences should share the strength or weakness of corresponding analogies.

4 The Virtues of Hypotheses

One way to get an overview of the virtues of hypotheses is to see them as spanning the ever-present gap between the universal aspiration and particular established theory and facts. We want a new hypothesis and a theory with great comprehension so that, if it is correct, it will have a maximum tendency to avoid future disappointments or disconfirmation. That is one of our methodological ideals. Still, we have contrasting ideals. Though we want generality, we also want testability, and connected with this is the ideal of preserving as much as possible of accepted theory, even as we go about changing it in light of contrary evidence.

No one would have even considered Einstein’s physics, without the assurance that it came up with the same predictions as Newtonian physics over the very wide range of circumstances in which Newtonian physics had succeeded in its predictions—so that even the boldest of hypotheses must have its conservative side. But, on the other hand, holding that every bold hypothesis must have its conservative side, does not plausibly amount to saying that new theories cannot be “revolutionary.”³¹ A new theory might plausibly be regarded as “revolutionary,” if it takes in new predictions, and preserves the evidence supporting its older competitor, while significantly modifying the principles or laws which allowed the comprehension of supporting evidence by the older competitor.

Seeing the virtues of hypotheses as spanning the tensions between the particularity of established fact and theory and ideal universality, as aiming us toward both predictive tests and general explanatory intelligibility, the other virtues fall somewhere between, in somewhat the following order: Refutability, conservatism, modesty, precision, elegance, generality. I want to suggest a continuum of the virtues with contrasting extreme points, approximately, from the virtues of the experimentalist to the virtues of the theoretician.

In Table 1, each has been outfitted with familiar, named excesses and defects, to help with their recognition in Aristotelian style, and the virtue of “*simplicity*” is understood as a component of both “*modesty*”—when accepted theory is chiefly retained—and of “*elegance*”—as involved in relating evidence and prediction to

³⁰ See, e.g., Hawking, Stephen (1988) [10] *A Brief History of Time*, pp. 105ff.

³¹ Contrast Barrow, John D. (2003) [2] *The Constants of Nature*, p. 60.

Table 1 Virtues of hypotheses

(Particular) ←-----//-----→ (Universal)

Excess: implausibility dogmatism meekness *Übergenauigkeit*^a naiveté rigidity

Virtue: Refutability Conservatism Modesty Precision Elegance Generality

Defect: self-insulation extravagance vanity vagueness complexity bias

(Experimentalist) ←-----//-----→ (Theoretician)

^a “*Übergenauigkeit*,” German: over-exactness; compare Latin, *meticulosus*: fearful

broader innovations of theory. (We tend to speak of “*elegance*” when a broader range of poignant evidence is comprehended by simplicity of law or principle.) The virtues of *refutability*, *conservatism* and *modesty* appeal to the experimentalist, because they involve only limited modifications of accepted theory, and because accepted theory tends to be built into the instruments and methods which have been used in testing and developing accepted theory. The further the theoretician departs from accepted theory, the more likely it is, in general terms, that innovative hypothesis and theory will make no clear predictions, even though they agree with all evidence so far established. In this direction science tends to the speculative. Still, it is not impossible for a very innovative theory to come up with plausible predictions of otherwise unforeseen phenomena, and a growing number of problems in accepted theory may bring the theoretician to the conviction that modest tinkering has degenerated into meekness, so that an entirely new approach to outstanding problems is required. *Precision* and *elegance* are particularly important in any broadly innovative approach, because *precision* (mathematical precision and quantification in particular) makes it more reasonable to expect a range of measurable results; and *elegance* holds the promise of a wealth of evidence comprehended on the basis of (relatively) simple principles—all of which would take us in the direction of greater *generality* than what was heretofore established.

If we follow Peirce in holding that “Logic may be defined as the science of the laws of the stable establishment of beliefs,”³² then we may doubt that there is, or could be, a *logic* of abduction which could establish stable beliefs regarding untested explanatory proposals on the basis of *laws*. Our list of the virtues of explanatory hypotheses is not a matter of general laws, since we have no general means of ranking the comparative importance of, say, conservatism and modesty or simplicity, precision or generality for an arbitrarily selected context of inquiry. In consequence, it seems we have no formal or law-like way to arrive at a generalized ranking of competing hypotheses which exhibit these virtues. The named virtues are

³² See Peirce, 1896 [16] [*Collected Papers*, Vol. 3, paragraph 429], where he continues by defining “exact logic”: “Then, exact logic will be that doctrine of the conditions of establishment of stable belief which rests upon perfectly undoubted observations and upon mathematical, that is, upon diagrammatical, or, iconic, thought.”

comparative terms. One hypothesis is judged to be more easily refutable or more conservative, more modest, simpler, more precise or more general than another hypothesis in relation to the same domain and context of inquiry, and the non-comparative, presumably monadic predicates, “refutable,” “conservative,” “modest” “simple,” etc. borrow what sense they have from the specific comparisons made in specific contexts. But even in a particular context of inquiry, when we are dealing with a specific domain and accepted theory and its problems, if we know exactly which of the proposed new explanatory hypotheses can be justly called more easily refutable, more conservative, simpler, etc., this alone does not tell us which hypothesis might best be accepted for preferential empirical examination in that context of inquiry. It is not that such judgments are not in fact made, and it is not that we cannot see the wisdom of examples. But our assembly of convincing examples of preferences among the virtues, from the contexts of particular inquiries, do not add up to a *logic*, as contrasted with an art, of abduction. In the context of possible theory change, we no longer know what to count as purely formal elements, as is perhaps evident, say, from Einstein’s revision of Newton’s definition of force, or from his revision of the concept of simultaneity.

On some occasions, generality overrules simplicity or conservatism, on other occasions modesty or conservatism rightly trumps generality or elegance. But if there were truly a *logic* of abduction, then we would expect general rules or some stable ordering of the virtues across distinctive domains, contexts and occasions of inquiry. Our esteem for any particular virtue in the order it provides to a range of hypotheses on a particular occasion appears to be bound to the particular context, and dependent on the specifics of content, and it is otherwise chiefly retrospective and *ex post facto*. No generalized ordering of explanatory hypotheses in terms of the virtues seems to be projectable across all domains and occasions of inquiry: not conservatism, not boldness.

This is not to say, however, that our intuitive sense for the value of particular analogies, and related hypotheses, cannot be improved by means of the study of the mathematics of model theory—which represents the abstract possibilities of projective or analogical mappings. A chief point, however, is that knowledge of the mathematical possibilities of mappings is empty without the knowledge of particularities of the various domains—which helps establish a required sense of relevancy and salience. Corresponding model-theoretic analogies are of general interest in supporting abductive inference, because model-theoretic analogies are intuitive or natural—at least for those who have some understanding of model theory—but also, in part, on the assumption that mathematics generally can be paraphrased into set theory. We might imagine, of course, that physicists, and other scientists, sometimes craft analogies supportive of particular hypotheses formulated in highly sophisticated mathematical terms, making no direct suggestion of model theory, strictly considered. But even in such cases, the proposals ought to be open to paraphrase into model-theoretic analogies.

Notice that physics seems lately to have discounted modesty and even refutability as it has explored the luxurious mathematical possibilities of string theory, drawn in this direction by the prospect of a unified quantum-mechanical theory of

the four known fundamental forces—a prospective theory which would bridge and reconcile quantum mechanics and general relativity. General relativity must break down, Stephen Hawking has argued, where the erstwhile continuities of curved space-time reach quantum-mechanical levels forbidding exact continuity of competing measurements.³³ Even at the time of Einstein’s early work, though it was known that Newtonian physics didn’t correctly predict the orbits of electrons around the atomic nucleus, the theory of relativity made only small and inadequate correction to these faulty predictions. In this context we understand the significance of Einstein’s role as the creator of relativity theory *and* as one of the chief thinkers responsible for the birth of quantum mechanics.³⁴ We see more clearly now, or believe more firmly, that continuity of space, time and motion must, in some fashion, give way to quantum-mechanical indeterminacies.

5 The Scientific Imagination

My conclusion is that in selecting among untested hypotheses, we have to do with a highly contextual type of judgment, a kind of art or wisdom arising from the expert’s extensive familiarity with the subject-matter. Selecting plausible hypotheses and weighing of analogies is not a formalizable skill so much as it is a matter arising from flexibility of mind in encompassing the details of a subject-matter. We have also found some reason to contrast the familiarity of the theoretician with that of the experimentalist. Deep familiarity with the details and problems of the domain of inquiry, on the part of a master of the discipline, is the one commonality which bridges those cases where we are inclined to favor bold generality and those where we may be inclined to favor more conservative, modest or easily testable alternatives. It is from this kind of perspective that we may judge of the lack of relevant detail, or the amateur status of wild guessing—as contrasted with educated guesswork.

The scientific imagination uses an organic classification, we might say, understood as a matter of strong, detailed analogies, while amateur fancy joins by accidental resemblances. The distinction parallels that between a more compelling argument from analogy and the kind of weak or false analogy which ignores detailed differences to focus on superficial similarities. In knowing what to count as potentially useful or more useful theoretical innovation and what might count as vain fancy, we depend on detailed reference to the particular subject-matter in the

³³ Cf. Hawking, Stephen (2006) [11] *A Briefer History of Time*: “...we know that the theory of general relativity must be modified. Because the classical (i.e. non quantum-mechanical) version predicts points of infinite density—singularities—it prognosticates its own failure...” See, p. 119 in the German edition.

³⁴ Einstein received the 1921 Nobel Prize in Physics, for his photoelectric law and work in the field of theoretical physics, thus chiefly for his work in the origin of quantum theory. Relativity, still under debate, was not mentioned.

continuity of inquiry. In initial evaluation of hypotheses proposed, we must start from a detailed and systematic account of past accomplishments in a field, together with the outstanding anomalies and problems. It is out of this tension that the properly disciplined and genuinely creative scientific imagination arises.

References

1. Aliseda, A.: *Abductive Reasoning, Logical Investigations into Discovery and Explanation*. Springer, Berlin (2006)
2. Barrow, J.D.: *The Constants of Nature*. Vintage Books, London (2003)
3. Callaway, H.G.: A role for Peirce's categories? In: Callaway, H.G. (ed.) *Meaning without Analyticity, Essays on Logic, Language and Meaning*, pp. 183–192. Cambridge Scholars Publishing, New Castle (2008)
4. Cao, T.Y. (ed.): *Conceptual Foundations of Quantum Field Theory*. Cambridge University Press, Cambridge (1999)
5. Einstein, A.: Approximate integration of the field equations of gravitation (1916). Translated in Engel, A. (ed.) *The Collected Papers of Albert Einstein*, vol. 6. Princeton University Press, Princeton (1997)
6. Feynman, R.: *QED: The Strange Theory of Light and Matter*. Princeton University Press, Princeton (1985)
7. Gabbay, D.M., Woods, J.: *The Reach of Abduction, A Practical Logic of Cognitive Systems*, 2 vols. North-Holland, Amsterdam (2005)
8. Gentner, D., Holyoak, K., et al. (eds.): *The Analogical Mind, Perspectives from Cognitive Science*. MIT Press, Cambridge (2001)
9. Greene, B.: *The Elegant Universe*. Vintage Books, London (1999)
10. Hawking, S.: *A Brief History of Time, from the Big Bang to Black Holes*. Bantam Books, New York (1988)
11. Hawking, S.: *A Briefer History of Time, the German edition*. Rowohlt, Hamburg (2006)
12. Hawking, S., Penrose, R.: *The Nature of Space and Time*. Princeton University Press, Princeton (1996)
13. Heisenberg, W., Pauli, W.: Zur Quantenelektrodynamik der Wellenfelder. *Zeitschrift für Physik* **56**, 1–61 (1929)
14. Magnani, L.: *Abductive Cognition, The Eco-Cognitive Dimensions of Hypothetical Reasoning*. Springer, Heidelberg (2009)
15. Marshall, W.C., Simon, R., Penrose, R., et al.: Towards quantum superposition of a mirror. *Quant-ph/0210001v1*, Sept. 2002, pp. 1–5
16. Peirce, C.S.: *The Collected Papers of Charles Sanders Peirce*, vol. 3. Harvard University Press, Cambridge (1931–1935)
17. Peirce, C.S.: Pragmatism as the logic of abduction (1903). In: Houser, N., Kloesel, C. (eds.) *The Essential Peirce, Selected Philosophical Writings*, vol. 2 (1893–1913). Indiana University Press, Indianapolis (1998), pp. 226–241
18. Penrose, R.: On Gravity's Role in Quantum State Reduction. *Gen. Relat. Gravit.* **28**(5), 581–582 (1996)
19. Penrose, R.: *The Road to Reality, A Complete Guide to the Laws of the Universe*. Vintage Books, New York (2004)
20. Penrose, R., Mercer, P.: Quantum Computation, Entanglement and State Reduction. *Philos. Trans. Math. Phys. Eng. Sci.* **356**(1743), 1927–1939 (1998)
21. Popper, K.: *The Poverty of Historicism*. Ark, London (1957)
22. Popper, K.: *The Open Universe*. Routledge, London (1982)
23. Popper, K.: *Objective Knowledge, revised edition*. Oxford University Press, Oxford (1979)

24. Quine, W.V., Ullian, J.S.: *The Web of Belief*, 2nd ed. Random House, New York (1978)
25. Rickles, D.: *Companion to Contemporary Philosophy of Physics*. Ashgate, Burlington (2008)
26. Rovelli, C.: 'Localization' in quantum field theory. In: Cao, T.Y. (ed.) *Conceptual Foundations of Quantum Field Theory*, pp. 207–232. Cambridge University Press, Cambridge (1999)
27. Smolin, L.: *The Trouble with Physics*. Houghton Mifflin, Boston (2006)
28. Stachel, J.: Introduction: quantum field theory and space-time (1999). In Cao, T.Y. (ed.) *Conceptual Foundations of Quantum Field Theory*, pp. 171–173. Cambridge University Press, Cambridge (1999)
29. Stachel, J.: Comments. In: Cao, T.Y. (ed.) *Conceptual Foundations of Quantum Field Theory*. Cambridge University Press, Cambridge (1999)
30. Thagard, P., Shelley, C.: Emotional analogies and analogical inference (2001). In: Gentner, D., et al. (eds.) *The Analogical Mind*. MIT Press, Cambridge (2001)
31. Woit, P.: *Not Even Wrong, The Failure of String Theory and the Continuing Challenge to Unify the Laws of Physics*. Random House, London (2007)