INTRODUCTION

Tamara Horowitz & Gerald J. Massey

The line between science and philosophy is sometimes drawn at observation. Observation itself can be passive and even unplanned, or active and artfully contrived. Observation of this second, deliberate sort is commonly known as *experimentation*. But when one reflects that scientific experiments are at least as likely to be *thought experiments* (*Gedankenexperimente*) as real ones, even the seemingly hard-headed appeal to observation to demarcate science from philosophy begins to look fanciful. Why? Because philosophers conduct thought experiments, too. Indeed, Rescher argues that the characteristic method of the ancient Greek nature philosophers (Thales and company) was the method of thought experimentation. And one of us (Massey) argues that thought experimentation has come to supplant meaning analysis as the distinctive method of contemporary analytic philosophy. Similarly, J. N. Mohanty contends that the distinctive phenomenological technique, the method of imaginative or eidetic variation, is essentially the method of thought experimentation. If claims like these foregoing ones are true, then an examination of thought experimentation goes to the heart of both science and philosophy.

Despite their centrality and importance to both science and philosophy, relatively little has been written about thought experiments. Ernst Mach wrote about them, but it seems that what has sparked contemporary interest has been a recent paper by Thomas Kuhn on the role of thought experiments in science. Kuhn concerns himself largely with what we will call the *paradox of thought experiments*, i.e., the initially puzzling fact that thought experiments often have novel empirical import even though they are conducted entirely inside one's head, so to speak.

Kuhn's resolution of the paradox of thought experiments harks back to Carnap's theory of bilateral reduction sentences. Carnap once insisted that new concepts enter science only by the route of explicit definition, an insistence that guarantees their complete eliminability. This definitional demand was soon recognized to be too restrictive, and Carnap relaxed it by insisting only that new concepts be introduced via an ensemble of bilateral reduction sentences of the form "$(x)\ [Sx \supset (Nx \equiv Rx)]$," where N is the new concept to be introduced and where S and R are already available scientific concepts. For example, N might be "magnetized," S be "placed near iron filings," and R be "draws the iron filings to it." Then, the bilateral reduction sentence entitles one to apply the predicate "magnetized" to an object if and only if it attracts iron filings when placed in their vicinity.

Bilateral reduction sentences have much of the "feel" of explicit definitions. However, a (bilateral) reduction sentence for a concept N is weaker

than an explicit definition of *N* in that the former gives only a sufficient condition for something's being *N*. But several reduction sentences for *N*, although they fail to furnish a single condition that is both necessary and sufficient for the application of the term *N* which they introduce, are collectively stronger than an explicit definition of *N*, in a sense now to be explained. An explicit definition is non-creative; its addition to a set of postulates yields a conservative extension (i.e., no new theorems couched exclusively in the old vocabulary can be derived). By contrast, the addition to a set of postulates of two or more reduction sentences for *N* is creative; new theorems couched exclusively in the old concepts can be deduced from the original postulates in conjunction with the reduction sentences.

Kuhn appeals to the just-mentioned creative trait of the way concepts are introduced in science to resolve the paradox of thought experiments, i.e., to explain how a method that restricts itself to analysis of concepts, and to known or presumed facts, laws, and theories, can lead to empirical predictions at odds with those very facts, laws, or theories. Kuhn's analysis helps also to explain why thought experiments proliferate in times of scientific crisis and during those anomalous episodes known as scientific revolutions.

A brief overview of the contributions to the present volume may give the reader a better picture of the range and significance of this important instrument of scientific and philosophical inquiry.

SECTION 1 – THOUGHT EXPERIMENTS IN THE HISTORY OF SCIENCE AND PHILOSOPHY

Nicholas Rescher

Rescher credits the presocratic nature philosophers with devising the technique of thought experimentation in order to explore *factual* reality through reasoning. Not surprisingly, then, Rescher takes an expansive view of thought experiments, virtually identifying thought experimentation with hypothetical reasoning. He then goes on to offer a six-cell taxonomy of thought experiments rooted in the arguments advanced by the presocratic nature philosophers, to wit: explanatory, refutatory, *reductio*, sceptical, analogical, and value-dominant thought experiments.

Rescher's classificatory scheme, the basis of which is as much structural as functional, should serve as an effective counterweight to the apparently widespread view that thought experiments are all of a piece, both as to structure as well as to function. Indeed, the reader will note that some of the papers in this volume come precariously close to advancing this suspect monolithic conception of thought experiments. So, the fact that Rescher's taxonomy is "imperfect," in the sense that its cells are neither mutually

exclusive nor jointly exhaustive, does not prevent it from playing an important prophylactic role.

Rescher's remark that Heraclitus was "deeply persuaded that it is ultimately by mind (which can contemplate what is not) rather than by vision (which can only contemplate what is) that the deepest truths are learned" suggests that the path from Thales to Kuhn is shorter than one might have supposed. In his closing remarks, Rescher appeals to Mach's maxim that real experiments should be preceded by thought experiments in order to defend philosophy against the charge that unbridled use of thought experiments renders it sterilely speculative. Far from stifling appeals to experience, Mach and Rescher believe that thought experimentation not only encourages such appeals but also marshals them in empirically fruitful and innovative ways.

Peter King

Pierre Duhem argued very persuasively that the distinctive features and elements of modern physics are already found or at least anticipated in medieval physics. But against Duhem, John Murdoch has recently emphasized the special *secundum imaginationem* aspect of medieval science. Peter King, too, directly challenges Duhem's claim, arguing that far from being *empirical* in any modern sense, medieval physics (more precisely, 14th century physics) was *academic* and *scholastic* in the pejorative senses of these labels. What in medieval physics sometimes look like direct appeals to experience turn out to be academic appeals to what is generally believed or thought to be known or the like.

Not content simply to challenge Duhem's claim, King articulates a conception of medieval scientific method rival to the one advanced by Duhem. By King's lights, the method of medieval physics is thought experimentation, and the methodology of medieval science is given by the medieval theories of *obligationes*. Unlike both Mach and Rescher, King denies that thought experiments, at least the medieval ones, foster or encourage real experiments. In medieval science thought experiments supplant real experiments. Rather than lead to real experiments, medieval thought experiments make the latter seem irrelevant and unnecessary. Nevertheless, King believes that thought experiments were well attuned to the role given them in medieval science, viz., to expose inconsistencies and conceptual inadequacies, to explore *apriori* truths, and to lend rigor and sophistication to the investigation of the necessary truths about nature.

For King, medieval thought experiments are arguments of either a hypothetical or counterfactual sort. Certain initial assumptions, often of an idealized or counterfactual character, are made, and the reasoner sets about to see what follows from them. To their everlasting credit, medieval thinkers recognized that thought experiments pose a peculiar and difficult

problem, namely, *what governs what happens in a thought experiment*? Or, to frame the problem differently, what rules determine what initial assumptions are permissible, what auxiliary assumptions can be introduced, and what follows from the set of initial and auxiliary assumptions? King argues that the theory of *obligationes*, the rules that specify what may and what may not be done in a medieval *disputatio* (a highly stylized form of academic debate), govern what happens in a thought experiment. As an example of an *obligatio* King cites Burleigh's rule. It obliges disputants to treat a sentence that is logically independent of the initial and auxiliary assumptions in the same way they should treat it in everyday rational life, i.e., to concede (deny) it if it is known to be true (false) and otherwise to suspend judgment. A medieval theory of *obligationes*, therefore, amounts to a theory of scientific methodology, where the method of science is thought experimentation.

In King's view, medieval science is conceptually very sophisticated, often impressively mathematical, but never empirical. Its success was judged by its ability to explain stock examples and to resolve counterexamples, not by its ability to predict new phenomena or to control nature. (He conjectures that Galileo was probably closer to the medievals in this respect than contemporary historians depict him to be.) King's claims are bold and well argued. Nevertheless, some thinkers will find it difficult to dismiss as nonempirical such appeals to experience as Buridan's invocation of a rotating blacksmith's wheel or of a spinning child's top as objections to Aristotle's theory of violent motion. They will be hard put to imagine what a more empirically grounded objection to Aristotle's theory would be like. But however this may be, there can be no doubt that King's paper presents a formidable challenge to certain well-entrenched views about the character of medieval science and about the respects in which it prefigured modern science.

Lilli Alanen

In the transition from medieval to modern philosophy no figure looms larger than Descartes. And if Lilli Alanen is right, the claim that Descartes was a key transitional figure extends to the theory of modalities no less than to epistemology and the philosophy of mind. Descartes himself conducted numerous thought experiments, quite as many as any medieval or contemporary philosopher. But his often-ingenious thought experiments are not the focus of Alanen's paper. Rather, she concerns herself with a related topic: the theory of possibility and necessity that underlies Descartes' bewildering doctrine of God's free creation of the eternal truths.

We remarked above that the medievals recognized the need for explicit rules that govern what happens in a thought experiment. We saw, too, that Peter King identifies these rules with *obligationes*, the prescriptions that regulate the give-and-take of medieval debate. But what happens when thought experiments transcend the bounds of the conceivable, or of the

possible? Do all such thought experiments degenerate into unintelligibility? Or do they, at best, just become *reductio* arguments? If not, what rules govern what happens when thought experimenters venture into the realms of the inconceivable and the impossible? Can philosophers chart these murky alien waters?

Unlike both his medieval predecessors and many contemporary philosophers, Descartes refused to identify the conceivable with the possible. Granted, he held that whatever we can (clearly and distinctly) conceive is possible, in the sense that God can bring about, exactly as we conceive it, whatever we clearly and distinctly conceive, if He so wills. But Descartes flatly denied that the fact that we cannot (clearly and distinctly) conceive of something means that it is impossible. He deemed it blasphemous to suggest otherwise, for then the power of God would be bounded by the finite human mind! For example, we cannot conceive how the sum of 2 and 2 could be different from 4. Yet we know that God could have made it so, because nothing escapes His infinite power, not even so-called *eternal truths*. Here, then, is but one of innumerable cases where the inconceivable is possible.

Alanen reviews a number of recent attempts to unravel Descartes' doctrine of the free creation of the eternal truths, and finds them all wanting, some more so than others. The exegetical-philosophical problem is to do three things simultaneously, to wit: make the doctrine *intelligible* to contemporary philosophers, *plausible* in its historical context, and *faithful* to Descartes' radical enterprise. It is relatively easy to accomplish any two of these goals in isolation from the third. What has proved to be exceedingly difficult is the joint realization of all three.

Alanen suggests that we will grasp Descartes' doctrine better when we understand what it is that he rejects. According to her, Descartes rejects all realist foundations for modality, such as those of Aquinas, Scotus, and Suarez. (We pause to note that Alanen and others credit Aquinas with a straightforward realist conception of possibility, perhaps unaware that in his much neglected tract *De Aeternitate Mundi contra Murmurantes*, Aquinas is driven by an argument of Algazel seriously to contemplate the view that God can do impossible things.) Alanen likens Descartes' "voluntarist" or "constructivist" theory of modality to that of George Henrik von Wright who locates necessity in an attitude taken toward propositions, but with this difference: von Wright acknowledges that we can and do change our attitude when it suits our purposes; Descartes thinks the attitude is woven into the fabric of our mind by the Creator Himself.

In conducting their thought experiments, philosophers often trespass beyond the boundaries of logical possibility. For example, they sometimes invoke worlds where the laws of logic are different from commonly accepted ones. The rocks, shoals, and reefs that ring Descartes' doctrine that the inconceivable, and even the logically inconsistent, may nevertheless be possible (for God) should serve as a somber warning of the hazards to be encountered

by metaphysical mariners who dare to sail these uncharted, and perhaps unchartable, waters.

SECTION 2 – THOUGHT EXPERIMENTS IN LOGIC AND MATHEMATICS

Dionysios Anapolitanos

Convinced that in mathematics the only experiments are thought experiments, Anapolitanos emphasizes the need for an adequate theory of mathematical *Gedankenexperimente*, the first prerequisite of which is a serviceable taxonomy. He foresees a four-stage theoretical enterprise: (1) a round up of thought experiments driven by a minimalist conception of what they are; (2) an initial sorting of the thought experiments thus collected into intuitively discernible kinds; (3) an analysis of how the recognized kinds of thought experiments function in the different fields and subfields of mathematics; and (4) identification of the conceivability conditions or constraints that bear upon the different kinds of thought experiments.

Anapolitanos himself lumps mathematical thought experiments into six somewhat overlapping groups. First, there are thought experiments that take place in as yet theoretically unstructured contexts. These experiments typically aim at answering such definite mathematical questions as what relations obtain among the numbers of the vertices, edges, and faces of regular polyhedrons. Second, there are thought experiments performed against the backdrop of a developed mathematical theory. Sometimes these thought experiments occasion modification or even abandonment of the original framework, e.g., certain thought experiments about the axiom of choice led to non-Cantorean set theory. Third, there are comparatively free-wheeling thought experiments conducted either during or in the immediate wake of a foundational crisis. Familiar examples are the thought experiments of Russell, Zermelo, Fraenkel, Gödel, Bernays and others which were directed at rehabilitating Cantor's naive notion of set that had been discredited by the antinomies. Fourth, goaded by disagreeable ignorance over the derivability or underivability from given axioms of a proposition or propositions that have come to be viewed as pivotal, some mathematicians undertake thought experiments aimed at resolving the matter. The most celebrated examples here, of course, are the thought experiments undertaken to settle the status of Euclid's parallel postulate. Fifth, philosophical ruminations—sometimes of a radical or revolutionary nature—can give rise to large-scale thought experiments, e.g., Brouwer's intuitionist program or Abraham Robinson's resurrection of infinitesimals. Sixth and last, there are thought experiments undertaken in order to secure a more hospitable conceptual framework for a particular field or subfield, e.g., infinitary combinatorics.

When he turns to the conceivability constraints on mathematical thought

experiments, Anapolitanos announces his preference for an itemized list over an attempted definition, explaining that any serious attempt at definition must presuppose such a list. On his own list, then, we find such items as the following: simplicity conditions (after all, mathematics was until quite recently an exclusively human enterprise); familiarity constraints (which cause mathematicians to eschew Henry Sheffer's stroke function in Boolean algebra or Gerald Massey's dyadic star function in closure algebra in favor of a multiplicity of more familiar and humanly tractable functions); plausibility constraints (a conservative element in mathematics, these constraints facilitate communication among mathematicians); efficiency constraints (these constraints are presently in flux because of the participation by computers in mathematical activity); reconceptualization conditions or constraints (here Descartes's algebraization of geometry and Gödel's arithmetization of syntax stand as sovereign examples); philosophical constraints (such as Hilbert's formalist philosophy of mathematics); and constraints of specificity and generality (for example, although arithmetic is not fully axiomatizable, mathematicians still require a characterization of the natural numbers determinate and specific enough for them to carry on the traditional work of mathematics).

Reflecting upon his own survey of mathematical *Gedankenexperimente*, Anapolitanos is justly moved to remark upon the rich variety of mathematical thought experiments and upon their indispensability in every nook and cranny of creative mathematics.

Barbara Massey

Gottlob Frege believed that a logic genuinely rival to one's own is inconceivable. Hardly deficient in imagination, Frege had no trouble conceiving a tribe that used an alternative logic. What he claimed to be unable to do was to picture the natives as *rational* when their logical rules conflicted with those he accepted. For, resolute Kantian that he was, Frege could conceive of a rule of logic as a *law* only if it was binding on all rational creatures.

As Barbara Massey points out, claims about alternative logics or deviant logics are often backed up by thought experiments that purport to show the possibility, or as in Frege's case the impossibility, of tribes that employ these logics. So, the question of whether there can be deviant logics commonly reduces to the question whether certain thought experiments are successful (rationally cogent) or not, i.e., whether certain imaginary scenarios are conceivable or inconceivable.

Conceivability and possibility are generally thought to be closely related if not altogether the same thing. We remarked above that Descartes is exceptional among philosophers in espousing only the first of the following two familiar metaphysical principles:

(i) *Whatever is conceivable is logically possible.*

(ii) *Whatever is inconceivable is logically impossible.*

Jointly, these two principles entail that conceivability and logical possibility coincide. Many philosophers—for example, David Hume who examines the relationship between conceivability and possibility at much greater length than does Descartes—readily accept both principles and employ them liberally in their philosophical work, as Hume does in the *Treatise* and first *Enquiry*.

What fascinates philosophers about conceivability is the free and easy access to possibility it seems to provide. And if the two aforementioned principles are true, conceivability really does furnish facile access to possibility. The ease of access derives from the allegedly privileged and immediate access we enjoy to the contents of our minds, among which are found our conceptions. If you are conceiving some state of affairs p, you have only to reflect or mentally notice that you are doing so. Then, merely by invoking the first of the above principles, you can be sure that p is possible.

Similarly, the attraction of inconceivability lies in the ready access it provides to impossibility. And again, if the second principle is true, inconceivability really does furnish a royal route to impossibility. For, thanks to your privileged and immediate access to your own mental powers, you can tell quite effortlessly that you cannot conceive some state of affairs q, i.e., that q is inconceivable. So, by appealing to the above two principles, you can be altogether confident that q is impossible.

Even if it didn't all sound much too pat and rosy, the *standard account* of conceivability and inconceivability, which is what we will call the foregoing pair of metaphysical principles, is hard to square with intellectual history. Frequently, what was at one time deemed conceivable or inconceivable is deemed the opposite at another. Even worse, the standard account warps mathematics beyond recognition. For, on the assumption that mathematical truths are logically necessary, to solve some open mathematical question Q, you need only to engage in the gentle exercise of trying to conceive Q. If you conceive Q, you know that Q is possible, and so true. Similarly, if you cannot conceive Q, you know that Q is impossible, and so false. Here, then, is a trivialization of mathematical discovery and justification in which nobody seems willing to acquiesce.

This brings us back again to the medieval question about what rules govern thought experimentation. The standard account of conceivability and inconceivability is at once plausible—after all, most philosophers embrace it—and indefensible, for the reasons we have just given. Massey herself flatly rejects the *standard account* of conceivability and inconceivability, which makes the rational cogency of thought experiments a matter of facile introspection. Further, she denies that there is any formal recipe for assessing the rational cogency of deviant-logic thought experiments, i.e., for determining whether one has succeeded in conceiving of a logically deviant tribe. She thinks that

there is no way to prove that logically deviant tribes are possible; one can only convince others of such a possibility by plausible story-telling.

In Massey's view, successful thought experiments are well-told stories, cogent because they have been fleshed out in compelling detail. So long as conception remains sufficiently schematic, virtually any state of affairs can be envisioned, even quite impossible ones. For example, it is easy to conceive a Turing-machine program that picks out the valid wffs of classical first-order logic—so long as one doesn't have to flesh this scenario out very much. But the impossibility of this state of affairs manifests itself in the insuperable obstacles that arise when one tries to flesh out its schematic conception. Therefore, it is precisely the fine-grained articulation of an envisioned scenario, the provision of relevant and abundant detail, that has the power rationally to convince people of the possibility of some conjured-up state of affairs. A good thought experiment is just a good, detailed story.

Massey examines several thought experiments that have been proposed to show the possibility, or the impossibility, of deviant logics. She finds all of them too schematic to carry rational conviction. She then undertakes to repeat Frege's thought experiment, with the intention of fleshing it out fully enough to see whether a logically deviant tribe is possible. Much of the paper is given over to her story of the Xenophobes, a people with an alien culture which they are determined to preserve, and an alien logic.

Massey tries to describe the Xenophobes in enough detail to make plausible her diagnosis of logical deviance. The leading idea behind her thought experiment, itself an exercise in anthropological story-telling, is value pluralism. Massey is convinced that the advocates of a privileged logic recognize only one logical value, usually truth. What primarily makes an inference rule logical for these theorists is its propensity to preserve that one value (truth). But if there are many logical values and if they cannot all be achieved simultaneously, the door is open to logical pluralism. One society might choose one value as supreme, and another society might choose another, and each might develop a logic appropriate to its choice. Massey's Xenophobes, for example, embrace values very different from those embraced by Frege, and it shows in their logical practice as well as in their logical theory. In good instrumentalist fashion, Massey argues that there is no rational way to subordinate some of these logical values to others in a way that would lead to a privileged logic. There are many possible logics, because there are many logical values.

SECTION 3 – THOUGHT EXPERIMENTS IN THE SCIENCES

Allen Janis

Philosophy of science and history of science have been erected, in large part, on the successes of science. It is likely that a more balanced under-

standing and appreciation of science, and possibly a deeper one, would have resulted if philosophers and historians had given equal consideration to the seamy side of science, i.e., to its failures. For example, philosophers have viewed scientific method as a battery of related canons that guarantee, or at least promote, scientific success. By contrast, one of us (Massey) has conjectured that, far from being oriented toward success, *scientific method is a miscellany of recipes that protect against various known kinds of scientific failure.* For example, the injunction to repeat experiments serves as one safeguard against the second kind of experimental failure enumerated by Janis (See below). It is both refreshing and enlightening, then, to find a philosophically sophisticated physicist, Allen Janis, concentrating his attention on thought experimental failure.

Having distinguished three principal ways in which real experiments can fail, Janis claims that thought experiments can fail in the same three ways. First, real experiments fail because they are not, or sometimes even cannot, be carried to completion. Reports of abortive undertakings end up in wastebaskets rather than in journals, so one seldom hears about these experimental failures. Second, real experiments fail by giving incorrect results for any of an indefinite number of reasons ranging from equipment failure to faulty experimental design. Third, real experiments fail when their results, albeit correct, do not decide the question or issue they were intended to resolve. Of these three kinds of failure, Janis finds the third not only the most intriguing but also the most important, for they sometimes lead to a deeper understanding of the situation that occasioned them in the first place.

Thought experiments can fail in the first way, e.g., when the thought experimenter lacks the knowledge or ability to carry the analysis of the imaginary scenario to completion. They can fail in the second way, too, e.g., when the thought experimenter makes a mistake or overlooks a relevant variable in his or her analysis. To illustrate such failure, Janis cites a thought experiment, proposed by Einstein at the 1930 Solvay Conference, in which an imaginary device seems to permit measurements that contradict the version of the Heisenberg uncertainty principle that involves time and energy. Bohr later showed that, ironically, Einstein had failed to take general relativity properly into account. Thought experiments can fail in the third way as well, i.e., by leading to correct conclusions that leave open or unresolved the question meant to be resolved by the thought experiment. An excellent example of such a thought experiment, we believe, is Newcomb's Problem which is discussed in Horowitz's paper (commented upon below) in this volume.

James Robert Brown

Brown distinguishes three functionally different kinds of thought experiments. First, there are *destructive* thought experiments that serve as a *reductio* of a theory, e.g., Schrödinger's cat which Brown takes to debunk the

Copenhagen interpretation of quantum mechanics. Second, there are *constructive* thought experiments that serve to support or even establish a theory, e.g., Newton's rotating waterbucket which Brown takes to confirm Newton's theory of absolute space and time. Finally, there are *platonic* thought experiments, *Gedankenexperimente* that are at once destructive and constructive, i.e., they simultaneously overturn one theory while establishing another.

Among platonic thought experiments Brown gives pride of place to the famous thought experiment wherein Galileo imagines what would happen if a heavy cannonball yoked to a light musketball were dropped from a height. Brown takes this *Gedankenexperimente* not only to overthrow the Aristotelian theory of freely falling bodies but also to provide apriori proof of Galileo's own theory that all bodies fall at the same rate (in a vacuum). But an even better platonic thought experiment, in our opinion, is the one Algazel advanced against Aristotelians who espouse an eternal-world thesis, i.e., who take the duration of the past to be infinite. Algazel asks us to calculate how many (human) souls would now be in existence if souls are individually immortal and if there have always been humans around and procreating at however modest a rate we like. One of us (Massey) has pointed out that Algazel's thought experiment drove even Thomas Aquinas to flirt in his *De Aeternitate Mundi contra Murmurantes* with the otherwise abhorrent proposition that God can do impossible things rather than to abandon the temporal-world thesis. The Church took the temporal-world thesis to be an Article of Faith, i. e., a revealed truth inaccessible to human reason. It follows from this, as Aquinas saw all too well, that Aristotle's eternal-world thesis had to be at least possible, something altogether ruled out by Algazel's ingenious thought experiment.

A less philosophically charged label for Brown's platonic *Gedankenexperimente* would have been *crucial* thought experiments, for like crucial real experiments Brown's platonic thought experiments purport to decide once and for all between competing theories. But Brown's heavily charged label is chosen with philosophical malice aforethought. A platonist like Gödel in mathematics, Brown thinks that we somehow perceive the eternal objects to which our mathematical terms appear to refer, thereby "explaining" how it is that mathematical truth lies within our ken. Also a platonist along the lines of David Armstrong about natural laws, Brown thinks that a well-conceived platonic thought experiment somehow enables us to see or perceive the law or laws of nature that constitute the theory established by the *Gedankenexperiment*. Brown's platonism about thought experiments is criticized at length by Norton in the latter's paper in this volume.

John Norton

Norton, like Brown, attempts to resolve what we have called "the paradox of thought experiments." However Norton, unlike Brown, does not endorse the view that thought experimentation involves a special non-empirical

way of knowing. Norton argues that if we take the position that a thought experiment is an argument of a certain sort we will be able to explain how thought experiments can provide us with new information about the empirical world. Further, we will be able to provide this explanation without assuming that thought experimentation requires us to engage in special non-empirical epistemic practices.

Norton first presents certain necessary conditions which an argument must meet in order to be a thought experiment. He then argues that a number of well-known thought experiments meet these conditions. Finally he explains how our understanding of these thought experiments is enhanced when their underlying argumentative structure is made explicit.

According to Norton, thought experiments are arguments which (i) posit hypothetical or counterfactual states of affairs, and (ii) invoke particulars irrelevant to the generality of the conclusion. The argument for (i) is that without the hypothetical or counterfactual element they wouldn't be thought experiments, but actual experiments, or descriptions of actual states of affairs. The argument for (ii) is that without the invocation of particulars which are irrelevant to the generality of the conclusion there would be no experimental appearance. Beyond these minimal requirements we find great flexibility. The arguments can be either inductive or deductive, and they can be carried out either wholly within some physical theory, or within some physical theory augmented by general philosophical principles.

It is central to Norton's account that the postulation of the particular must in principle be eliminable from the argument and that therefore the argument doesn't *really* depend on the particular which is invoked, (although it may not be easy to follow the argument once the particular is eliminated). In fact, Norton argues that thought experiments can be classified according to how the particulars are removed. In cases of type 1, which are usually *reductio* arguments, the particular-free conclusion follows deductively from those premises which do not involve the postulation of the particular. In cases of type 2, the conclusion is freed of the particulars by an inductive step, such as an assertion that the case involving the particular is "typical" or that the particulars are "inessential."

Examples of thought experiments of type 1 can be found in thermodynamics, since the three laws of thermodynamics can easily be stated as assertions that certain situations or "machines" are impossible. Thus to derive a theorem X, one can take one of these laws as premise, assume not-X, and show that not-X allows the design of a machine disallowed by the premise. Norton argues that one of Einstein's arguments for the wave-particle duality of black body radiation is a good illustration of an argument of this form.

Examples of thought experiments of type 2 can also be found in Einstein's work. Norton reconstructs certain of Einstein's thought experiments to

make their argumentative structure explicit. According to Norton's reconstruction, certain philosophical theses about verifiability appear along with the relevant principles of physical theory as premises. Norton argues that although Einstein didn't state these philosophical principles explicitly his arguments partially rest upon them. In particular, a careful reconstruction of Einstein's arguments show that they rely on one or the other of the following principles. Either (1) a theory should not use theoretical terms which have no observational support, or (2) states of affairs which are not observationally distinct should not be distinguished by the theory.

Finally, Norton argues that when the argumentative form of a thought experiment is made explicit, it will be clear whether the argument is deductive or inductive. If the argument is inductive the inductive step will be readily identifiable. and whether the argumentative form is deductive or inductive, all implicit philosophical assumptions likewise will be readily identifiable. Once all of this is done we will be able to apply all of the usual standards used to evaluate arguments or philosophical principles. These standards of evaluation are part and parcel of our usual philosophical and scientific practice. Thus thought experimentation does not require some kind of "mysterious new window on the physical world."

Andrew Irvine

Irvine responds to the view of Rescher that the presocratic philosophers devised the technique of thought experimentation. Although Irvine agrees that the presocratics introduced the use of hypothetical reasoning, *reductios*, and explanatory conjectures, into the study of nature, he argues that to engage in thought experimentation is to do more than engage in these forms of reasoning. He argues that the presocratics are better understood as having taken a first step toward the eventual development of thought experimentation.

As his first example of a presocratic use of the method of explanatory conjecture Irvine discusses Thales' conjecture that the earth is like a floating log. Thales made this conjecture to explain the stability of the earth's position. The conjecture is supported because it explains the original confirmed belief. Irvine asks if this example shows that Thales has anticipated the modern theory of abduction, according to which hypotheses are selected and evaluated on the basis of their explanatory power. He answers this question affirmatively, though it is a guarded "yes." He then asks whether Thales' reasoning was a full-blooded thought experiment. To this question his answer is negative. He claims that Thales' reasoning was no more than an argument from analogy. It was not imbedded in a detailed theoretical context.

Another example which Irvine discusses is Euclid's proof that there exist infinitely many primes. He finds this a more promising example, since it

did arise in a well-developed theoretical context. Nevertheless, since the subject matter is mathematics, Irvine doesn't count this as an early use of thought experiments in investigating the *physical* world.

Irvine goes on to address the following question: If, as he has argued, it is not the case that just any instance of hypothetical reasoning ought to count as a thought experiment, what further conditions should an instance of hypothetical reasoning meet in order to qualify as a thought experiment? He argues that the answer to this question will differ from science to science. In each branch of science the conditions that an argument must meet in order to be a thought experiment in that branch of science will depend on the conditions that an experiment which is not a thought experiment must meet in that branch of science. Since experiments in astronomy differ in some ways from experiments in the other sciences, Irvine investigates a thought experiment in astronomy. He discusses Olbers' paradox according to which it follows from certain allegedly plausible assumptions that the sky should appear as if it were covered by one great sun.

According to Irvine, the lesson to be learned from this example and others, is that a thought experiment, as opposed to a mere instance of hypothetical reasoning, must stand in a special relationship to the science of which it is a part. Features of this special relationship include the following: the argument must be relevant to the testing of some hypothesis which has arisen in that science; many of the assumptions made in the course of the argument must be supported by empirical observation occurring in the pursuit of the special science; the thought experiment must be set out in enough detail; and the outcome of the thought experiment should be relevant to accepted theory of the special science.

Since Irvine, like Norton, considers a thought experiment to be an argument, we will mention two points where they differ. First, Irvine thinks that it is *not* necessary for thought experiments to have a counterfactual element, since many thought experiments are meant to precede real experiments. Second, Irvine characterizes thought experiments as taking place wholly within some particular scientific theory, whereas Norton explicitly allows the inclusion of philosophical principles among the premises of at least some of the thought experiments he examines.

Ronald Laymon

Like Brown and Norton, Laymon is primarily concerned with the question of how thought experiments, which proceed by postulating unreal situations, can provide us with information about the physical world. Laymon answers this question differently from both Brown and Norton. He argues that the fact that thought experiments depend on the consideration of idealized situations does not affect their relevance to the real world

as long as a special relationship exists between a certain set of real, or "truly possible," experiments and the thought experiment. His account of the special relationship is essentially as follows:

Suppose there is a thought experiment containing an argument from premises $(\exists x)Tx$ and $P_1 \ldots P_n$, to a conclusion Q, where $(\exists x)Tx$ is a highly idealized experimental description, the P_i are laws or principles believed true, and Q is to be demonstrated. Laymon claims that many thought experiments contain arguments of this form. Often the argument is unsound, since

$(\exists x)Tx$, being highly idealized, is in fact false.

Either of two approaches may be available. First, there may be transformations of the unsound argument into other arguments which form a series, each member of which represents the content of a "truly possible" experiment where the analog of $(\exists x)Tx$ is true, and which series approaches the given argument in the sense that the analogs of T approach T as an asymptote. Second, we may have theoretical accounts of relevant disturbing forces (e.g. friction), which accounts generate either a series of truly possible experiments asymptotically approaching the thought experiment as disturbing factors are minimized, or a series of truly possible experiments asymptotically approaching the thought experiment as progressively more refined theories of the disturbing factors are applied. Laymon illustrates this analysis with thought experiments due to Stevin, Mach, and Gouy.

Laymon considers the possible objection to his view that if one looks to the history of science one doesn't find scientists attempting to justify their use of thought experiments by constructing, in principle or in fact, the appropriately related series of real experiments. Nevertheless they seem confident in their use of thought experiments. Laymon suggests that thought experiments have been convincing in the absence of such argumentation because when an informed audience is considering a thought experiment it will naturally come to see the analogy between the thought experiment and real experiments which asymptotically approach the thought experiment in the appropriate way. Laymon argues further that where thought experiments have not been successful at convincing their audiences of the intended conclusion it is because of the absence of perceived analogies with existing experimental and analytic work.

Mark Wilson

Wilson claims that Kuhn's approach to the question how thought experiments can contribute to the advancement of science fails to provide a way to understand the contributions to physics of an important class of thought experiments.

According to Wilson, Kuhn holds that the purpose of thought experiments must either be to clear up conceptual confusion or to bring to our attention the fact our theories do not fit well with some of our background

data. But Wilson argues that this account of the function of thought experiments won't suffice for all thought experiments—not all thought experiments show a theory to have been either conceptually muddled or empirically inadequate.

The particular example Wilson explores in some depth arose in the course of the work of d'Alembert and Euler on the theory of the vibrating string. Although they agreed on the proper equation for the system, they disagreed as to what should happen, according to their equation, in a thought experiment in which a string is released from a plucked initial position—that is, one with a triangular starting configuration. In a modern version of the thought experiment, we can imagine an infinitely long string plucked at a point, so that the string's initial displacement is triangular in shape. Idealizing in sensible ways with respect to real-string damping and other factors, the appropriate differential equation of the string's motion appears to require that at every point the operative forces are proportional to a derivative which, unfortunately, is undefined at the triangle's peak.

Wilson takes us through a discussion of d'Alembert's and Euler's responses to this mismatch between the analytic description of a physical process and its geometrical description. He also discusses a range of other possible responses to it, and several other similar problems. He turns at last to what he calls "the most common reaction to these problems," which is to maintain that it is a mistake to formulate physical laws in terms of differential equations. Instead one should begin with integral forms. As he shows, this approach avoids the mismatch between analytic and geometrical descriptions confronted by a differential equation approach.

But, he goes on to claim, the move to integral forms leaves us without enough equations to set a determinate, well-posed physical problem in the case of the nonlinear string. As it happens, he claims, thermodynamics provides the missing condition needed to set things right—in particular, the Clausius-Duhem inequality. Wilson points out that it seems that one can conclude that classical continuum mechanics lacks the closure one might naively expect. Thus, Wilson claims, the plucked-string thought experiment leads to a conceptual advance which does not fall neatly into either of the categories envisaged by Kuhn.

John Forge

Forge claims that what distinguishes physical science from other branches of science is that the aim of physical science is to investigate the relations among quantities that are found in the world. Given the importance of the concept of "quantity" for physics it is interesting to note that many of the investigations of this concept rely on the postulation of imaginary situations. Forge raises the question whether the discussions that involve these

imaginary situations should be considered as involving thought experiments, and, if so, whether they are good thought experiments.

All of the hypothetical examples which Forge discusses are designed to answer the question whether quantities are properties or whether quantities are classes. The first example he takes up is due to Brian Ellis. Ellis considers a universe with just one object which Forge calls "e." The question is whether such an object can have quantitative properties. If it cannot, perhaps this shows that quantities have an essentially relational character.

Forge asks whether Ellis's use of this hypothetical situation counts as performing a thought experiment, and, if so, whether it is a good thought experiment. Forge argues that Ellis's example meets the conditions which Norton suggests a thought experiment must meet. Nevertheless, Forge doesn't think the thought experiment is completely successful. He claims that this is because the counterfactual situation posited is *too* different from the actual world. Forge suggests that for a thought experiment to be successful the condition ought to be added that the counterfactual situation envisaged sufficiently resemble situations that actually exist. With this suggestion in hand Forge evaluates other arguments for or against an extentionalist view of quantities.

Finally, Forge argues that since it is harder to tell when his "sufficient resemblance" condition is met in philosophy than it is in science due to the greater development of scientific theories, judgments about the effectiveness of thought experiments in philosophy will be more difficult to make than judgments about the effectiveness of thought experiments in the sciences.

James G. Lennox

Lennox suggests a very different solution to the paradox of thought experiments. He argues, in effect, that the problem that the other authors have been addressing sometimes doesn't exist, since sometimes the role of thought experiments is not to demonstrate that a theory is true or false, but rather is to reveal the theory's explanatory potential. He argues that this is the important role that thought experiments played in Darwin's *On the Origin of Species*.

Lennox presents his case by first explaining the relationship between Darwin's work and that of Charles Lyell. There were two important aspects of Lyell's work that are relevant here. First he described the agencies of geological change as working gradually, and second he hypothesized only causal mechanisms presently in operation rather than relying on the intervention of factors to which humans have no epistemic access. According to Lennox, Darwin adopted this methodology, with one important difference. Lyell collected data on the actual operation of the mechanisms, such as erosion, that he invoked in his explanations, even if this data

reflected only the operation of these mechanisms in the short run. In place of this sort of data gathering, Darwin presented *imaginary* illustrations.

What is the value of Darwin's imaginary illustrations? According to Lennox their point is to show that *if* each of the processes postulated by his theory somehow interacted, then there *could* occur an accumulation of minute, random variations in a particular direction culminating in a new species. That is, their point is to show, contrary to the objections of some of his contemporaries, that his theory *could* explain the coming into existence of a new species.

Lennox claims that, interestingly, thought experiments are still the appropriate form of argument to use when the very possibility of a Darwinian explanation for some phenomenon is at issue, rather than when an actual Darwinian explanation is at issue. The fact that today's thought experiments rely on computer models makes them no less thought experiments than Darwin's own.

Sarah Thomason

In an essay teeming with fascinating examples from her own research, Sarah Thomason distinguishes two types of thought experiments commonly found in linguistics: explicitly theory-governed *Gedankenexperimente* that, though themselves non-probative, point toward or suggest real experiments or empirical data that actually support or undermine a given theory or hypothesis, and introspective *Gedankenexperimente* that themselves qualify as evidence for or against a given theory or hypothesis. The first type of thought experiment usually bears upon universals of language structure, language learning, language change, or language use; the second type, upon features of particular languages.

Thomason believes that the first type of *Gedankenexperimente* resembles the thought experiments found in physics and other sciences. In linguistics as in these sciences, the challenge is to find or to orchestrate real-life analogues to the imaginary scenarios, ones similar enough to carry probative force. Consider, for example, the warmly debated question whether one language can so influence another that a bonafide *mixed* language results, i.e., a language descended from two parents. Debate on this issue was hardly idle, because a presupposition of the Comparative Method in linguistics rules out a language's having more than one parent.

In order to address the mixed-language hypothesis, Thomason conducted a thought experiment in which speakers of English retain their grammar intact but borrow their vocabulary wholesale from Russian. Thomason recognized that no linguist was going to be convinced of the possibility of mixed languages by this thought experiment alone, for skeptics would demand an account of the historical linguistic processes that can lead to such a mixture. Rather, what her thought experiment did was to suggest

where to look for actual examples of reasonable facsimiles of mixed languages. Acting on this hint, Thomason soon found some, most notably a Tanzanian language *Ma'a* whose vocabulary is mostly Cushitic and whose grammar is almost entirely Bantu. Curiously, *Ma'a* arose over three centuries in the direction opposite to the one stipulated in the thought experiment: Cushitic speakers borrowed Bantu grammar while keeping their own vocabulary.

The second type of linguistic *Gedankenexperimente*, the introspective kind that does not point to or suggest real-life probative analogues but rather itself serves as evidence for or against a given hypothesis, often concern the appropriateness of a given linguistic form or construction. These introspective thought experiments are notoriously subject to the experimenter effect (experimental bias), i.e., the experimenters' theoretical expectations cause them to find just what they had expected to find. Thomason approvingly cites Labov's findings that, whereas Chomsky and Chomsky-influenced graduate students judged "We received plans to kill me" to be grammatical and "We received plans to kill each other" ungrammatical—exactly as one of Chomsky's theories predicted— the judgments of subjects who were not linguists ran exactly opposite.

In syntax, testing hypotheses against introspective thought experiments is the rule rather than the exception, so methodological safeguards against the experimenter effect become important. Drawing on riveting examples, Thomason shows that the experimenter effect is even more pronounced when the introspecting agent is an informant rather than the linguist. Unfortunately, Thomason sees no way to circumvent introspective thought experiments in certain parts of linguistics, in syntax in particular. For, unlike in other sciences, in parts of linguistics the primary data for or against a theory or hypothesis lie inside people's heads.

SECTION 4 – THOUGHT EXPERIMENTS IN PHILOSOPHY

J. N. Mohanty

We have seen that Rescher takes an expansive view of thought experimentation, virtually identifying it with hypothetical reasoning. By contrast, Mohanty thinks that the concept of a thought experiment becomes significant only when the notion is restricted to imaginative representations of processes or operations that cannot be performed outside the mind. Mach's mental rehearsals of real scientific experiments, therefore, would not qualify as thought experiments in Mohanty's sense. On the other hand, the *Critique of Pure Reason* turns out to be a treasure trove of genuine thought experiments. For example, Kant invites his readers to join him in mentally stripping away from an instance of empirical knowledge everything due to the understanding and its concepts so as to isolate an empirical intuition. Then, from this

empirical intuition, Kant asks them mentally to purge everything due to sensation so as to isolate pure intuitions of space and time. What makes these mental exercises thought experiments is the fact that nothing, in particular no amount of mental gymnastics, could ever result in someone's actually having an empirical intuition uncontaminated by the understanding, or in someone's actually having a pure intuition of space and time.

Mohanty carefully explains the phenomenological method known as imaginative variation or eidetic variation and shows it to be a species of thought experimentation in his restricted sense of the term. He debunks some exaggerated claims made in behalf of this method by Husserl and others, namely, that it is apodictic (yielding necessary truths) and incorrigible. But he defends the method of eidetic variation against the twin charges of circularity (you must already know what the essence of X is in order to decide whether an imagined case qualifies as an X) and psychologism (the method reduces objective possibility to subjective imaginability). In both cases, Mohanty's defense begins with concession: the method is circular, and, unlike possibility, imaginability is subjective. But the circularity is the virtuous one of *meaning clarification* (which, we note, is much like Carnapian *explication*), and possibility is not reduced to imaginability but only explored thereby.

The medievals asked what governs what happens in a thought experiment. Mohanty asks what governs what happens, and what should happen, when one begins to imagine variations on some arbitrarily chosen exemplar of an X. What determines whether one should regard the variation as lying outside, or within, the range of the concept X? Mohanty's answers to these questions depend on his construal of the method of imaginative variation as a corrigible technique for uncovering meaning, and as a useful device for getting at eidetic possibility (which is stronger than logical possibility) through imaginability.

Mohanty recognizes that in recent times analytic philosophers have taken up thought experimentation, in his narrow sense, with a passion. Their mania for thought experimentation roughly coincides with the wholesale invocation of possible worlds in recent analytic philosophy. Mohanty himself rejects many analytic thought experiments because they are based on mere logical possibility, not on eidetic possibility. From the perspective of phenomenology, a thought experiment in which a person P splits into two persons with both of whom P is represented as not only physically (bodily) continuous but strictly identical reveals nothing about the concept of a person. Only if such a scenario could be imagined, i.e., intuited in some tough sense of the term and therefore eidetically possible, would it bear on the meaning of "person" or, in the jargon of classical phenomenology, on the essence of a person.

Mohanty seems to locate the phenomenological method somewhere between the analytic method (the method of contemporary analytic philoso-

phy) and the scientific method. The scientific essence (alternatively, the scientific meaning) differs importantly from the phenomenological essence (phenomenological meaning). The former is recondite, hidden behind or beneath the phenomena, and so is inaccessible to armchair cogitation. The latter is the invariant law of the phenomena, and so lies open to the armchair method of imaginative variation based on eidetic possibility, but not to the armchair method of freewheeling analytic philosophy which is based on logical possibility.

Mohanty's essay raises an interesting question with which he does not himself deal: Do scientists *qua* scientists ever conduct thought experiments (in Mohanty's restricted sense of the term)? Doubtless many scientific *Gedankenexperimente* are mental representations of real experiments, i.e., experiments that could actually be carried out, at least when science has advanced to a certain stage. But some scientific thought experiments, indeed some of the most important ones in the history of science, seem to be unrealizable in this way. Do these experiments bear on phenomenological essences rather than scientific essences, or on the former more than on the latter? Many analytic philosophers claim there is no sharp line between science and philosophy. Should the foregoing considerations goad phenomenologists to subscribe to this same thesis?

Rolf George

George claims that a new and still influential style of epistemology, which he labels *epistemological thought experimentation*, became fashionable in the 18th century. This novel approach to epistemology is constructivist: attempts at imaginative reconstruction of the cognitive process supplant the definitional and criteriological concerns of earlier epistemologists.

George identifies three stages in the imaginative reconstruction of the cognitive process. First, there is an attempt to characterize the initial states of the perceiver or knower. Not uncommonly, these initial states are taken to be non-intentional states akin to sensations of pain or pleasure. Second, there is a sustained effort to imagine or construct mental capacities and mental processes of a sort that, when brought to bear on the cognitively impoverished initial states, yield such cognitively rich products of the mind as perceptions of external objects or causal beliefs about the world. Third and finally, sweeping philosophical lessons or morals are drawn from successes and failures of the thought experiments conducted during the second stage. Typically, these morals assume the form of what have come to be regarded as philosophical positions, e.g., Humean empiricism, transcendental idealism, and common-sense philosophy.

Condillac, Berkeley, Hume, and Kant were leading practitioners of epis-

temological thought experimentation, but the movement hardly died with the sage of Königsberg. On the contrary, George traces this 18th century brand of epistemological constructivism through Carnap's *Aufbau* and Goodman's *Structure of Appearance* right up to current work in Artificial Intelligence and cognitive science.

While conceding that the tradition of epistemological thought experimentation, like other traditions in philosophy, is an artifact of the historian, George claims that this tradition, like the others, is not a creation *ex nihilo* but an interpretive taxonomy with a solid *fundamentum in re*. It is an illuminating fact about 20th century foundationalist epistemology, for example, that it stands well outside this constructivist tradition, despite certain superficial traits that suggest its inclusion.

One should not suppose that constructivist thought experimentation constitutes only one strand in the epistemological endeavors of its 18th century practitioners, or an inconsequential one. Very little would be left of Hume's or Kant's systems of philosophy, George believes, if their epistemological thought experiments were excised. By contrast, by conceptualizing modern philosophical constructivism so that it falls within the tradition of epistemological thought experimentation, George highlights what is perhaps most important about and most central to this recent and contemporary work, and what marks its continuity with ongoing work in Artificial Intelligence and cognitive science.

Gerald Massey

Mohanty and Massey agree that recent analytic philosophy is rife with thought experiments, and that these conceptual scenarios are predicated on logical possibility rather than on some more robust modal notion. But Massey contends that so-called Modern Philosophy already teemed with thought experiments of the same uncritical sort found in contemporary analytic philosophy. What is uncritical about them is implicit in what Massey calls the *thesis of facile conception*. This thesis is closely related to what above we called the *standard account* of conceivability. It permits the invocation of casually alleged conceivability to establish possibility, and of casually alleged inconceivability to demonstrate impossibility.

According to Massey, the success of recent science posed a serious threat to philosophy because recent science imperialistically claimed the entire cognitive terrain for itself. In a desperate attempt to retain a cognitive niche, analytic philosophers latched onto meaning analysis which promised an extra-scientific route to truths. And because these truths were both necessary and apriori, they even seemed to give philosophy a cognitive upperhand over science. But when meaning analysis was discredited through the debunking, by Quine and others, of the analytic-synthetic distinction,

analytic philosophers were left bereft of any distinctive method that would individuate them from scientists, a development that pleased a few philosophers as much as it unsettled most.

But analytic philosophy did not shrivel up and die. According to Massey, at the very moment analytic philosophers were throwing analyticity out of the front door of their discipline they were conspiratorially letting it re-enter the backdoor in the guise of conceivability. Hence, what took the place of meaning analyses were thought experiments or, as Massey sometimes calls them, *conceivability arguments*. But whereas analytic philosophers had developed stringent standards of meaning analysis, almost no one seemed to appreciate that, without suitably rigorous standards of conceivability, thought experiments were moot. Unlike their medieval precursors, analytic philosophers did not pause to ask what governs what happens in a thought experiment. And unlike such modern precursors as David Hume (when philosophizing at his best), they did not conduct their thought experiments against the backdrop of an articulated theory of conceivability.

Massey believes that serious theories of conceivability, and appropriate conceivability standards, are likely to be developed only insofar as they become relevant to the march of science. Hermann von Helmholtz's elaboration of a sophisticated theory of spatial conception, one grounded in then-contemporary physiology, to show that non-euclidean geometries are conceivable is taken by Massey to confirm his claim. Another confirmation comes from the reaction of logicians and mathematicians when a claim like Church's thesis is challenged by a thought experiment or conceivability argument. They will accept as evidence bearing on Church's thesis only mental scenarios that meet certain strict, albeit largely implicit, standards of conceivability. Off-the-cuff conceivability claims, the sort of claims that analytic philosophers bandy about in their dialectical interchanges, are accorded absolutely no credit by logicians and mathematicians when these claims purport to bear on serious scientific business like Church's Thesis. What is good enough for philosophy is seldom good enough for science.

Two obvious morals emerge from Massey's paper. First, philosophers should muster enough intellectual integrity to eschew conceivability arguments that fail to measure up to the standards that have been developed piecemeal in science. Second, philosophers ought not rest content with piecemeal developments but should instead turn their talents to the construction of a general theory of conceivability.

Richard Gale

According to Gale, not only do analytical philosophers offer *analyses* of concepts, i.e., necessary and sufficient conditions for the correct application of the concept in all possible circumstances, but they also mercilessly

subject their analyses to test through thought experiments. Gale thinks that often a thought experiment designed to test an analysis of a given concept is taken to show the inadequacy of the analysis of the concept when in fact it reveals something quite different, namely, empirical presuppositions of the language game in which the concept is embedded. Gale calls these thought experiments *perverse*. But what is perverse about them is not the thought experiments themselves, but rather how they are taken to bear on the analyses for which they were conjured up. Shorn of such perversity, these thought experiments play the useful and sometimes important role of unmasking the empirical presuppositions of the relevant language games.

Gale advances a primitive taxonomy of thought experiments that he deems useful to the enterprise of turning perverse thought experiments into useful ones. He bundles together in his first taxonomic cell those thought experiments that yield clear-cut counterexamples to a given analysis of a concept. Gale's own example of an imagined shaven rooster as a clear-cut counterexample to the proverbial analysis of *human being* as featherless biped suggests that clear-cut counterexamples are not themselves clear cut. After all, a shaven rooster remains a feathered animal no less than a quadruple-amputee war veteran remains a biped.

The second cell of Gale's taxonomy houses thought experiments that generate undecidable cases, i.e., cases where the rules of use of the concept under test neither determine that it applies, nor determine that it does not apply. These thought experiments expose the inescapable penumbra of vagueness that surrounds any analyzed concept. However, Gale thinks that many undecidable-case thought experiments are mistaken for clear-cut-counterexample ones, especially when the concept in question is multi-criterial.

In a genuine undecidable-case thought experiment, it matters crucially whether the undecidable case is common or rare. If it is common, to play the language game in which the concept is embedded is futile, for empirical presuppositions of the associated language game are false. But if the undecidable case is rare, the language game retains its value, as well as its point, undiminished. (We note here an affinity between Gale's treatment of undecidable-case counterexamples and Hume's theory of general rules. Hume devised his theory to justify the differential handling of rare versus common counterexamples to general claims, such as the singular missing-shade-of-blue counterexample to his general maxim that all simple ideas are copies of simple impressions.)

Gale thinks that the science-fiction thought experiments that philosophers notoriously direct against analyses of personal-identity concepts are typically perverse. Contrary to their customary acceptation, these thought experiments show, not the inadequacy of the analyses, but the falsity *in the imagined world* of the presuppositions of the personal-identity lan-

guage game. In such a world, this language game would have neither point nor value. Wittgenstein himself advanced such thought experiments, but without succumbing to the aforementioned perversion. He saw clearly that what the counterexample discredited was not the analysis of the concept but the presupposition that the associated language game made sense in the envisioned world.

Gale looks closely at the language game of personal identity, claiming that it is an empirical presupposition of this game that persons be unique in ways that have human importance, i.e., that they not be interchangeable *salva aestimatione*. We diverge to make an important point. Do the racist and the non-racist play the same personal-identity language game? On Gale's account, they do not. The members of other races are value interchangeable for the racist, so his person concepts do not apply beyond the provincial boundaries of his own kind. Similarly, the exploitation of animals is predicated upon a supposed value interchangeability of members of a given animal species, an interchangeability which is no more justified in fact than is the value interchangeability of the members of other races.

Tamara Horowitz

Horowitz notes that it is a standard feature of the methodology of decision theory to test the deliverances of putative principles of rational choice against intuitive judgments concerning which courses of action open to imaginary agents in thought experiments are rational. One notorious such *Gedankenexperiment* is Newcomb's Problem. Some theorists claim Newcomb's Problem favors causal decision theory with its two-box strategy over evidential decision theory (sometimes called Bayesian decision theory) with its one-box strategy, while others claim Newcomb's thought experiment is too farfetched to have any bearing on such issues. Horowitz argues that both parties are right, but also that both are wrong.

Horowitz claims that to be relevant to decision theory a thought experiment must incorporate numerous normative epistemic elements. There are two principal ways to insert these normative elements into a *Gedankenexperiment*: implicitly through detailed description of the sort at which novelists excel, and explicitly through abstract stipulation of the sort philosophers commonly employ. A thought experiment presented via stipulation is called a *schema* to suggest its unformed or indeterminate character. Such a schema is *realized* or made determinate by a detailed description that is consistent with it. Herein lies the rub. Often there is no way to realize or flesh out a schema that (1) comports with all its normative stipulations while (2) retaining its probative relevance. Newcomb's Problem, a schema in Horowitz's sense, is defective in just

this way: no realization of the Newcomb schema is relevant to the theory of rational choice.

Horowitz does not claim that there are no realizations of Newcomb's schema. Quite the contrary. She herself presents a realization of the schema that features the Gullibles, an imaginary people who differ epistemically from us in important ways. In particular, when they carefully sift through the evidence imagined in Newcomb's problem, Gullibles find a Great Predictor hypothesis, as opposed to a Great Hustler hypothesis, highly credible. So, compatibly with causal decision theory and the dominance principle, the Gullibles adopt a two-box strategy, in apparent opposition to evidential decision theory. But the apparent rejection of evidential decision theory by the Gullibles in no way counts against evidential principles dictating *our* rational choices. Rationality is one thing for Gullibles, quite another for us.

Horowitz also examines a quasi-realization of the Newcomb schema in which causal and evidential decision theory both dictate a modified one-box strategy. We say "quasi-realization" because not all the normative elements of Newcomb's original problem are preserved. Consequently, the convergence in this scenario of the competing decision theories holds little interest for the theory of rationality.

Several years ago Richard M. Hare made similar points about the thought experiments typically advanced as counterexamples to utilitarianism. According to Hare, these allegedly decisive thought experiments—all of them schemata in Horowitz's sense—appear decisive only when left schematic. Once fleshed out, they are found either to be self-inconsistent or else not to discredit utilitarianism at all. In a sense, then, Horowitz develops and generalizes the line of argument used by Hare to defend utilitarian moral theory, and applies the articulated generalization to decision theory. But there is an even earlier anticipation of this line of argumentation. It is not farfetched, we believe, to interpret Hume's indictment of miracles as an argument to the effect that no realization of any miracle schema supports the religious hypothesis.

Stephen Hetherington

When commenting above on Barbara Massey's paper, we remarked that conceivability and possibility are generally thought to come pretty much to the same thing, as are inconceivability and impossibility. We remarked too that Hume, like most philosophers, accepts both of the following two metaphysical principles:

(i) Whatever is conceivable is logically possible.

(ii) Whatever is inconceivable is logically impossible.

Jointly, these two metaphysical principles entail that conceivability and logical possibility coincide, as do inconceivability and logical impossibility. We called this view the *standard account* of the matter.

We remarked also that what fascinates philosophers about conceivability and inconceivability is the free and easy access to possibility and impossibility they seem to provide. And if the two aforementioned metaphysical principles are true, they really do furnish facile access to these modal features of reality. The ease of access derives from the allegedly privileged and immediate access we enjoy to the contents of our minds, among which are found our conceptions.

Hetherington concerns himself, not with the ontological principles (i) and (ii), but with their evidential counterparts (i*) and (ii*):

(1*) If you can conceive *p*, then you are justified in believing that *p* is possible.

(2*) If you are unable to conceive *p*, then you are justified in believing that *p* is impossible.

Hetherington raises the question whether possibility (impossibility) and conceivability (inconceivability) are linked together only evidentially, or whether one can infer the latter from the former and so obtain modal knowledge from conceivability and/or inconceivability.

In a manner reminiscent of familiar knowledge-as-justified-true-belief-plus-something-else analyses, Hetherington carefully sifts the relations between modalities, representations, justification, belief, and knowledge. Because of such phenomena as Escher's paradoxical representations of the impossible, Hetherington takes especial pains to distinguish representation as actual from representation as possible and to see what bearing this distinction has on his principal concern, to wit: the relationship between conceivability and inconceivability on the one hand and modal knowledge on the other. Along the way Hetherington raises such fascinating questions as whether Hume's empiricist system has room for the distinction between representation as actual and representation as merely possible.

The standard account confers parity on conceivability and inconceivability, recognizing them both as royal routes to modality, i.e., royal routes to possibility and impossibility, respectively. Hetherington's analysis never goes beyond evidential links, though it does so without preserving parity. On his analysis, for example, conceivability is more closely related to possibility than inconceivability is related to impossibility.

Joseph Camp, Jr.

Whereas inferences to logical modalities from (in conceivability premises are notoriously treacherous, Camp argues that inferences from such premises to strong epistemic modalities are no less deceptive. He illustrates his

point with a parable about hunter-gatherers, one of whom (Mary) thinks she hears something making a snuff-snuff sound in a cottonwood grove. Dubbing the snuff-snuff source *Clyde*, Mary thinks Clyde is a large animal, probably a moose, but concedes under questioning that Clyde may be something quite different, perhaps even a feisty muskrat. But what Mary is certain of, and so refuses to give ground on, is that Clyde went snuff-snuff in the woods. After all, Clyde *must* have gone snuff-snuff in the woods, for this is the one property of Clyde that she cannot imagine or think away. One speculation leads to another and soon the tribe have constructed an elaborate story about a giant moose whose carryings-on neatly explain a host of their observations, including Mary's snuff-snuff experience. But eventually the tribe come to believe, for good reasons, that the moose Clyde never made it to the grove, notwithstanding Mary's certain belief that Clyde went snuff-snuff among the cottonwoods.

Camp claims that Mary's little thought experiment—Clyde *must* have gone snuff-snuff in the woods, for try as she might, Mary cannot think away this feature of Clyde without losing him as an object of thought altogether, and so she concludes that it is *absolutely certain* that Clyde went snuff- snuff in the woods—belongs to a genre of logically defective thought experiments that contains many important philosophical *Gedankenexperimente*. Camp characterizes the genre somewhat like this:

> An object of thought x is contemplated in a context so impoverished that some one feature F is necessary for retaining x as an object of thought. The thought-experimenter then concludes that x must be F, and goes on to interpret this to mean that it is absolutely certain that x is F.

As examples of this defective genre of *Gedankenexperimente*, Camp cites the following: Descartes' inability in the thought experiment of the *Second Meditation* to think away thinking as a feature of himself, thereupon concluding that he must think and so that it is absolutely certain that he is a thinking thing; Leverrier's inability in Saul Kripke's thought experiment about Neptune to think away the property of causing the perturbations in the orbit of Uranus, and so concluding that Neptune must cause the perturbations of Uranus's orbit and thus that it is certain (apriori) that Neptune causes them; and the inability of Cartesian-influenced empiricists, in thought-experiments about their seeming to feel cool, to think away the feature of their sensation being of coolness, and so concluding that it must be of coolness, thereby taking themselves to have established that if they seem to feel cool then it is certain that they feel cool. Camp suggests that a survey of the history of philosophy—from contemporary to ancient times—would uncover legions of thought experiments of this logically defective kind. As instructive as such a survey might be, Camp suggests that the important thing is for philosophers to avoid this peculiarly *Gedankenexperiment* fallacy in the future.