THOUGHT EXPERIMENTS
ROY A. SORENSEN

New York
OXFORD UNIVERSITY PRESS
1992
The Logical Structure of Thought Experiments

In the study that I am making of our behavior and motives, fabulous testimonies, provided they are possible, serve like true ones. Whether they have happened or no, in Paris or Rome, to John or Peter, telling impart useful information to me. I see it and profit from it just as well in shadow as in substance.

Montaigne

This chapter lays out a classification scheme for thought experiments. Don't worry about whether this is the uniquely correct scheme. The adequacy of a classification system is more a question of efficiency and suggestiveness. A good scheme consolidates knowledge in a way that minimizes the demand on your memory and expedites the acquisition of new knowledge by raising helpful leading questions. For example, one of the virtues of the upcoming taxonomy will be that it doubles as a theory of fallacy. The reason is that once one understands how each part of a thought experiment is organized toward the ultimate end of modal refutation, one can systematically diagnose failure. Admittedly, this analysis plays up the logical aspects of thought experiment at the expense of its psychological side; but the overrepresentation will be evened out in chapter 10.

I. Attributing Thought Experiments

Whenever we describe the mental life of another person (and even ourselves), we interpret. This may seem illegitimate; we ought to be discovering the person's ideas, not inventing them. But you can only unearth psychological facts if you dirty your hermeneutical shovel. Dream description poses the same problem. If the dream is simple, as when you dream you are eating a onion, then you can report it in good conscience—but probably won't because it's a bore. If the dream is complicated, as when you dream of hunting predatory onions beneath the surface of thirty-century Io, then your report must order the chaos. No justice can be done to the dreamy quality of the story because reports must impose categories familiar to both speaker and audience. The only way to avoid distortion is to avoid reporting the dream; but then, no interesting dreams are told. Although interpretative distortion is less severe when you are articulating your own beliefs, there is the lingering feeling that you are touching them up as you go along.

These problems intensify when we describe the beliefs and reasoning of other people. Interpretative intrusion is inevitable. So grin and bear it! Since the distorted version is usually of higher logical quality than the raw original, make an upbeat comparison; touched up beliefs are like computer enhancements of dull, blurry photographs.

Our interpretative machinations surface in a muted form with regular experiments. No experiment is completely public. The reason is that the identity of an experiment is partly determined by the experimenter's intentions. So two behaviorally equivalent experiments can be distinct. Otto von Guericke proved that sound travels through water by always ringing a bell before he fed fish in a pond. Eventually, hungry fish would come as soon as the bell rang. If this format were followed in order to demonstrate classical conditioning, it would be a psychological, rather than a physical experiment. It is tempting to say that the experimenter's intentions are always decisive in determining the nature of the experiment. But this would be a version of the intentional fallacy, which occurs when the artist's intentions are given decisive say over the nature of the artwork. The problem of experiment attribution is magnified in the case of thought experiments because the physical component is smaller or nonexistent. There are no experimental instruments to examine, no manipulations to witness, no subjects or substances to probe. Thought experiments are done in the head.

Thought experiment attribution is profitably compared to the attribution of arguments. The interpreter must first decide whether an argument is being propounded. Sometimes the speaker's limited stretch of discourse makes the matter plain. Otherwise, we must ask the speaker whether he is arguing for a point or merely asserting something. Unhappily, most people are unfamiliar with logical terminology and so are crude informants of their own reasoning. If we decide that the speaker is indeed arguing, then his logical ignorance will hamper our efforts to specify the argument.

Arguments are typically presented as enthymemes. An enthymeme is an abbreviated formulation of the argument, that is, a formulation missing a premise or conclusion. For instance, 'Only citizens can vote, so Carsten cannot vote' omits the premise 'Carsten is not a citizen'. Since the gain in brevity is constrained by the desire to be understood, there is a limit on how short an enthymeme can be. If we clip too much, the misshapen utterance becomes ambiguous or incomprehensible. Be mindful of the sloppiness that is bred by brevity. Logic helps us avoid this vice by giving us the ability to make arguments explicit. The premises of the regimented argument can be carefully inspected for plausibility and relevance. This naked state also provides the best basis for comparison with other arguments: hidden parts become visible, and false parts disappear. Thought experiments have lacked comparable quality
control. The upcoming classification system is intended to fill this regulative
gap by requiring structural details.

Admittedly, the unclarity of enthymemes and ellipsis is not always a matter
of remedial obscurity. Sometimes the problem is vagueness. I believe that men
are taller than women but I do not believe that all men are taller than all
women or merely that some men are taller than some women. I believe
something in-between but cannot exactly specify which proposition. A similar
vagueness dogs arguments. The trouble scholars have had in formulating
Wittgenstein’s private-language argument suggests that Wittgenstein may not
have had any exact argument in mind. Thought experiments can be equally
evasive.

Thought experiment attribution is further complicated by the principle of
charity. This principle tells us to maximize the rationality of interpretees
(choose interpretations that minimize the vices of bias, circularity, inconsis-
tency, etc.). Hence, when we attribute a thought experiment, we should try to
make it harmonize with other claims made by the thinker. For example, David
Hume’s “missing shade of blue” experiment appears to refute his principle that
basic ideas must be derived from simple impressions. To know how a pineapple
tastes, you must first taste it; to know what ‘purple’ means, you must first see
it—and so on for all ideas that are not built out of other ideas. Hume begins by
noting that shades of color are also simple, that they can resemble each other
and can be laid out in a gradation uniting quite different colors.

Suppose therefore a person to have enjoyed his sight for thirty years, and to
have become perfectly well acquainted with colours of all kinds, excepting one
particular shade of blue, for instance, which it never has been his fortune to
meet with. Let all the different shades of that colour, except that single one, be
placed before him, descending gradually from the deepest to the lightest; ’tis
plain, that he will perceive a blank, where the shade is wanting, and will be
sensible, that there is a greater distance in that place betwixt the contiguous
colours, than in any other. Now I ask, whether ’tis possible for him, from
his own imagination, to supply this deficiency, and raise up to himself the idea of
that particular shade, tho’ it had never been conveyed to him by his senses?

This thought experiment is readily classed with others aimed at the refutation
of general principles. In this case, the target is concept empiricism—the view
that all our simple ideas are obtained from experience. Hume himself is a
famous exponent of this doctrine. The mystery for historians is the cavalier
manner in which he dismisses this apparently deadly counterexample. Hume
blithely continues, “I believe there are few but will be of opinion that he can;
and this may serve as a proof, that the simple ideas are not always derived from
the correspondent impressions—tho’ the instance is so particular and singular,
that ’tis scarce worth our observing, and does not merit that for it alone we
should alter our general maxim.”? Perhaps we are entitled to ignore excep-
tional situations when designing a simple set of rules or when streamlining for
pedagogical efficiency. But Hume is laying down the foundation of his philo-
sophical system, an enterprise in which practical considerations do not justify
dismissal of rare counterexamples.

II. Thought Experiments as Alethic Refuters

Although there are a number of ways to classify thought experiments, a
refutational format scores the most points when judged by familiarity, specific-
ity, and simplicity. According to this scheme, thought experiments aim at
overturning statements by disproving one of their modal consequences. Modal-
ities are operators that are applied to propositions to yield new propositions.
There are deontic modalities (permissible, forbidden), epistemic modalities
(know, believe), and alethic modalities (possible, necessary). The alethic mo-
dalities are the best-understood and more-basic modality. Hence, we won’t miss
anything by concentrating on them.

There are two alethic operators. Necessity is symbolized with a box, □, and
works like the universal quantifier ‘all’. That is, when you say that p is neces-
sary, □p, you are saying that p is true in all possible worlds. An example is that
in all possible worlds, triplets have siblings. Possibility is symbolized with a
diamond, ◇, and works like the existential quantifier ‘some’. When you say
that p is possible, ◇p, you are saying that p holds in at least one possible world.
For example, in some possible world, moths outnumber spiders. The impossi-
bility of p is symbolized as ◇¬p and means that p is not true in any possible
world. For instance, in no possible world is there a bladeless ax missing a
handle.

Picture thought experiments as expeditions to possible worlds. The mission
is to refute a source statement that has an implication about the constituents
of these worlds. Sometimes the consequence is that p fails to hold in any possible
world. Hence, if we find a possible world in which p is true, we have refuted
the consequence and thereby the source statement. Likewise, if the source im
ples that p holds in some possible world and we show that it holds in none, then
the source is indirectly refuted by the refutation of ◇p.

A. Necessity Refuters

A necessity refuter is a supposition designed to refute a statement by showing
that something ruled out as impossible by that statement is really possible after
all. The necessity argument can be extracted from the following quintet of
jointly inconsistent propositions:

i. S Modal source statement. Fertile sources of modal propositions
include semantic theses (definitions, synonymy claims, entailment theses),
propositional theses (unverifiability, unfalsifiability, indetectability), feasi-
ability claims, law statements, disposition and intention attributions,
validity verdicts, and clusters of these—theories.
ii. \( \square I \) Modal extractor. This proposition draws the relevant modal implication from the source statement.

iii. \( (I \& C) \square W \) Counterfactual. This proposition is read as a subjunctive conditional: if I and C were the case, then W would be the case. This proposition claims that the antecedent, which is the conjunction of the implication and the imagined situation, has a weird consequence.

iv. \( \sim \diamond W \) Absurdity. This proposition explains the weirdness as an impossibility.

v. \( 
\square C \) Content possibility. This asserts that the content of the thought experiment is a possibility.

B. The Five Responses to the Quintet

Since the above five propositions are jointly inconsistent, one cannot hold all five. This means that there are at most five consistent responses to the set.

1. Bad source statement. The aim of a thought experiment is to refute the source statement, just as the aim of chess to checkmate your opponent. Although our intentions normally match the aims of our activities, we are sometimes driven by other goals (call these non-aim goals "motives"). A chessplayer's motive may be fame, the amusement of a friend, or the demonstration of an opening. Likewise, a conductor of thought experiments may have sundry ulterior motives for deploying his hypothetical case. We must nevertheless understand his moves as means to the end of refuting a targeted statement.

a. Definition busters. We can illustrate the process with a thought experiment that started a storm of philosophical analysis. In 1963 Edmund Gettier published a small discussion note entitled "Is Justified True Belief Knowledge?" His target was the "JTB" (justified true belief) definition of knowledge: A knows that p if and only if (i) A believes p, (ii) A is justified in believing p, and (iii) p is true. Nearly all epistemologists accepted this definition or a slight variant. Gettier's objection was that the definition is too broad. To prove the point he asked us to suppose that Smith and Jones are candidates for the same job. We are then to suppose that Smith acquires justified belief that

a. Jones is the man who will get the job, and Jones has ten coins in his pocket.

For example, Smith's evidence might be that the president of the company told Smith that Jones will get the job and that Smith has previously counted the coins in Jones's pocket. Smith notices that (a) entails

b. The man who will get the job has ten coins in his pocket.

and so comes to justifiably believe (b). Now comes the twist; despite what the president said, Smith is the man who will get the job and, as it happens, Smith has ten coins in his pocket. Hence, b is true, and Smith justifiably believes it. Yet Smith does not know b. Too much luck was involved. Thus, the JTB definition of knowledge is false.

Gettier's thought experiment can be recast into regimented form:

G1. The definition of knowledge is justified true belief.

G2. If knowledge is justified true belief, then necessarily, if a person has justified true belief that p, then he knows that p.

G3. If all justified true believers that p have knowledge that p and Smith is justifiably right but for the wrong reason, then Smith knows that (b) because of luck.

G4. It is impossible for anyone's knowledge to be due to luck.

G5. It is possible for Smith to be justifiably right for the wrong reason.

Almost all epistemologists agree that the Gettier cases are counterexamples to the JTB definition of knowledge and so reject the first member of the set. This was the response Gettier intended to elicit. A few commentators deny the possibility of a justified false belief and so deny the fifth proposition on the grounds that Smith was not justified in believing in (b).

1. Refuting full definitions. Reportive definitions, the kind dictionaries market, describe a pattern of usage and so are vulnerable to empirical refutation. If you can show that people do not really use "sweat" in the way Merriam-Webster defines it, then Merriam-Webster is mistaken. Reportive definitions can be refuted by sampling actual usage found in books, newspapers, and conversations. Occasionally, these observations are supplemented with experimental evidence in the form of interviews, discourse prompts, or questionnaires about accepted usage. But definition testers also make great use of hypothetical scenarios. Purely stipulative definitions cannot be proved false because they lack a truth value. If I declare that 'flem' means the food residue left on a knife, then I am laying down a rule, rather than describing a fact. But this does not make the stipulative definition immune to criticism; it just changes the criteria of criticism. We criticize flat, rules, and orders chiefly on grounds of inefficiency. The stipulator has some special goal. So my stipulation can be undermined by showing that it cannot achieve its goal or that its goal can be achieved in an easier or more reliable way. Stipulators are also beholden to background goals that create the potential for further efficiency critiques. Hence, definers are pragmatically committed to defending more than their definition. By bringing the definition to our attention, they suggest that it is worthy of our attention. But it is only worthy of our attention if it scores well on the standard criteria for grading definitions. Hence, the definer commits himself to the claim that the stipulation achieves a variety of objectives (or would achieve them if the definition were adopted). This metaclaim about the
efficacy of the definition does have a truth value and is open to proof or refutation.

Other definitions combine a reportive element and a stipulative element. Different sorts of definition have different purposes for the stipulation. Precisifying definitions, for example, are intended to reduce vagueness. If I offer a million-dollar prize for the biggest puppy, then a reportive definition of 'puppy' (such as 'A puppy is an immature dog') would cause disputes and confusion as to what counts as a puppy. Although the definition correctly reports ordinary usage, it leaves too much open to dispute along the borderline between immature and mature dogs. So for the purposes of my big puppy contest, I should define 'puppy' as, say, 'dog younger than one year'. Administrators of government subsidies have also learned the value of 'precisifying' definitions and so provide precise criteria for 'poor', 'citizen', and 'dependent'. 'Precisifiers' do not have carte blanche; they are obliged to confine most of their stipulation to the gray areas. Their definition of 'poor' cannot imply that millionaires are poor or imply that penniless mothers are not poor. Moreover, their stipulation must reduce vagueness. Defining 'poor person' as 'person without enough money to achieve an adequate standard of living' fails to eliminate borderline cases. So precisifying definitions have a stipulative element and a reportive commitment, and either aspect can be assessed by thought experiment.

The stipulative element of operational definitions is intended to increase the verifiability of the term. In the case of theoretical definitions, we aim at integrating the _definiendum_ with key terms of the theory and tightening its link with laws, tests, and constants. Defining 'acid' as a proton donor strengthens its relevance to established chemical and physical theory, thereby maximizing its informativeness. Defining 'sound argument' as 'a valid argument with true premises' links soundness with validity and thereby to the concept of logical form and generalizations about truth preservation. Legal definitions also aim at this integrative goal, so that users of the term are quickly directed to the relevant parts of the law.

Theoretical definitions are not restricted to science. Anywhere there is theory, there is theoretical definition. Consider theology. Here we find efforts to systematize vocabulary and eliminate conceptual problems with special definitions of 'omnipotence', 'sin', 'faith', and so on. Thought experiments abound. In attempting to define 'resurrection', Thomas Aquinas used the possibility of pure cannibals to eliminate the following definition: 'resurrection' means the reconstitution of a person from his own matter. Who would get the matter of a cannibal who ate nothing but people, the cannibal or his victims? Victims and cannibals must be resurrected for Judgment, yet there is not enough matter to go around. So Aquinas concluded that the definition was mistaken and recommended that we require at most the same _type_ of matter, rather than the same _particular_ matter.

ii. **Refuting partial definitions.** Thought experiments do not need sharply demarcated targets. They can be deployed against fuzzy, partial definitions and rough notions of what a word means—"folk semantics."

iii. **Refuting definitional policies.** Some doctrines provide a general framework for defining a whole family of terms. For instance, behaviorism says that
adequate definitions or psychological terms (hunger, hope, grief) are to be drawn in terms of the behavior of organisms to which those terms are applied. A more recent view is functionalism. According to this doctrine, mental vocabulary refers to functional states of the cognizers in question. A consequence of this view is that functionally equivalent systems have the same psychological attributes regardless of differences in what they are made of. Hence, functionalists say a computer would think if it shared your input/output relations.

John Searle attacks functionalism with his Chinese Room thought experiment. He has us imagine that he is in a room with a manual telling him how to write Chinese symbols in response to incoming Chinese symbols. The symbols are meaningless to Searle because he does not know how to read Chinese. However, the manual is in English and only requires that he recognize the shapes of the symbols and correlate them with other shapes. Searle blindly follows the manual and produces outgoing batches of symbols in response to incoming batches. Chinese speakers outside the room think they are having a conversation with Searle. They submit questions in Chinese and Searle writes back with “answers” that seem sensible and fluent to his audience. Despite appearances, Searle does not understand Chinese. The lesson is that purely syntactic competence does not suffice for comprehension. But note that computers will only be able to “communicate” with us by the sort of figure shuffling Searle simulates in the Chinese Room. Since functionalism implies that the symbol manipulation would be enough for understanding, Searle concludes that the doctrine is false.

iv. Scope expanders. Scope theses are nice examples of nondefinitional source statements. Scope propositions say that a principle only applies within a certain domain of phenomena. For example, the uncertainty principle is often assumed to operate only within the realm of elementary particles. However, there are various thought experiments designed to show how the uncertainty can creep into the behavior of middle-sized objects. Imagine two spheres each the size of a baseball. The first is glued to the floor of a vacuum chamber while the second is dropped on the first. Given perfect aim and the absence of air resistance, vibrations, and so on, the ball will bounce up and down indefinitely after striking the stationary sphere’s apex. But the uncertainty principle says that there is a chance that the ball will land a little bit off center. These tiny deviations will then accumulate with each bounce until the ball lands on the floor.

v. Enriched source statements. Necessity statements can also be derived from equivalence, implication, and incompatibility claims. Our discussion of definition and synonymy illustrates how thought experiments are used against semantic equivalence claims. But since the notion of necessity can be strengthened beyond logical necessity, nonsemantic equivalence theses should be recognized: “Water is H₂O.” “A straight line is the path of an undisturbed moving object.” “A species is an interbreeding group of organisms.” Just as logical necessity is compatibility with logical laws, physical possibility is compatibility with physical laws, biological possibility is compatibility with biological laws, and economic possibility is compatibility with economic laws.

Principles professing to state empirical laws can be refuted by thought experiment. This is done by interpreting the modal operators as governing natural necessities instead of merely logical ones. Hence, a biological principle can be refuted by showing that it entails biological impossibilities such as reverse evolution (forbidden by Dool’s law).

If we interpret the modal operators in terms of compatibility with moral laws, we can treat many of the famous hypothetical counterexamples of ethics. For example, some people defend their meat eating by appealing to the fact that human beings are superior to animals. Vegetarians counter that this defense is adequate only if it is necessary the case that any superior being is entitled to eat any inferior being. But if these were so and superior aliens landed, then they would be entitled to eat human beings for just the reasons we have (taste and convenience). But that would be inhumane! So since extraterrestrial gourmets are a possibility, vegetarians urge omnivores to recant the superior-being defense.

2. Misconnection. Since modal fallacies are common, expect some thought experiments to quickly entail. One modal muddle was a favorite target of medieval logicians: the confusion between a conditional being necessarily true—□ (p ⊃ q)—and its consequent being necessarily true—p ⊃ □ q. Sceptics commit this fallacy when they argue as follows:

1. If you know that p then you must correctly believe that p is true. Kp ⊃ □ (p & Bp)
2. It is possible for you to mistakenly believe that p. □ (¬ p & Bp)
3. You do not know that p. ¬ Kp

If we interpret the argument according to the symbolism on the right, the argument is valid, but the first premise is false. The premise is true if interpreted as giving the necessity operator large scope over the whole conditional: □ (Kp ⊃ □ (p & Bp)). But then the argument is invalid. The fallacy is also committed by some hard determinists. Their first premise is that if you perform an action, then you cannot do otherwise. They continue by saying that a man performs a free action only if he could have done otherwise and so conclude that no man is free. Here, the culprit is the ambiguity between the true but trivial conditional □ (Ap ⊃ Ap) (Necessarily, if you act to bring about p, then you act to bring about p) and the false but frightful conditional Ap ⊃ □ Ap.

“Logical” fatalists commit another modal fallacy, confusing the necessity of a disjunction, □ (p v ¬ p), with the necessity of its disjuncts, □ p v □ ¬ p. They note that it is necessarily the case that either I will surf tomorrow or not. But if it is necessarily true that I will surf tomorrow, I am not free to not surf. And if it is necessarily true that I will not surf tomorrow, then I am not free to surf.
Hence, either I am not free to surf or I am not free to not surf. Therefore I am fated one way or the other.

Sometimes the thought experiment is intended to show that a statistical “certainty” has been mistakenly identified as a physical necessity. In the nineteenth century, the basis for the second law of thermodynamics was Clausius’ axiom that it is impossible for heat to pass from a colder to a warmer body unless some other change accompanies the process. In 1871 James Clerk Maxwell suggested that a creature small enough to observe and handle individual molecules could violate the law. He introduced what has become known as “Maxwell’s demon”:  

If we conceive a being whose faculties are so sharpened that he can follow every molecule in its course, such a being, whose attributes are still as essentially finite as our own, would be able to do what is at present impossible to us. For we have seen that the molecules in a vessel full of air at uniform temperature are moving at velocities by no means uniform. . . . Now let us suppose that such a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole, so as to allow only the swifter molecules to pass from B to A. He will thus, without expenditure of work, raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics.  

Although Maxwell once said he wished “to pick a hole” in the second law, he did not believe the law to be false. The chief end of his creature was “to show that the 2nd Law of Thermodynamics has only a statistical certainty.” More picturesque is Maxwell’s famous verdict: “Moral: The 2nd law of thermodynamics has the same degree of truth as the statement that if you throw a tumblerful of water into the sea, you cannot get the same tumblerful of water out again.” Hence, although the thought experiment is aimed at the law, Maxwell aims to refute only the claim that the law states a physical, as opposed to a statistical, necessity.

Many economic laws have been criticized as having only statistical necessity. Consider the law stating that the quantity sold varies inversely with its price. We can show that it is not universally true with a thought experiment featuring an isolated village of peasants. The villagers can exist on a weekly ration of one pound of rice and two pounds of beans or can exist on the tastier and more nutritious fare of one pound of rice, one pound of beans, and one pound of fish. Rice and beans each cost $1.00 per pound, but a pound of fish costs $1.20. Since the peasants are paid $3.20 a week, they spend it all on the rice, beans, and fish diet. What happens if the price of beans rises to $1.10? (It is impossible to buy less than a pound of anything.) Bean sales double because the only adequate diet becomes the one with one pound of rice and two pounds of beans.

3. **Erroneous counterfactual.** We have already witnessed how the medieval recipe for making a vacuum by freezing a container full of water was ruined by a false counterfactual; that is, since cooling contracts most substances, the medieval naively inferred, “If the water froze, it would contract.”

There are subtler counterfactual errors. When Galileo realized that changes in motion were important, not uniform motion, he became more interested in acceleration. But there were two competing definitions: change of velocity with time and change of velocity with distance. The latter was favored by Leonardo da Vinci. Since Galileo believed that a body in free-fall has uniform acceleration, he tried to derive different empirical predictions from the definitions. This led him to an a priori argument against Leonardo’s definition. If the velocity of a body is proportional to its distance, then a peculiar race could be arranged. For suppose we roll two balls down the same inclined plane but stop one of them at the halfway mark. How much longer before the second ball reaches the finish line? Galileo’s answer: the second ball would arrive there instantaneously. At the halfway mark, the first ball has only half the velocity that the second has at the finish line. Since the second ball has twice the velocity, it therefore travels twice the distance and so is at the finishing line when the first ball is halfway to the finish line. Since the balls obviously would not behave this way, Galileo rejected the definition of velocity as change with distance. Although later physicists agreed with Galileo’s thesis that velocity is change of distance over time, they dismissed his thought experiment as fallacious; for it rests on a confusion between terminal velocity and average velocity. Leonardo is committed to saying that the ball that has traveled twice as far has twice the velocity, but he is not committed to saying that the ball immediately acquires the twofold speed. Leonardo is only saying that the ball will eventually increase up to double the speed, not that it will average double the speed throughout its trip.

At the Solvay conference in Brussels in 1930, Einstein presented a thought experiment that was intended to refute a version of Heisenberg’s uncertainty principle, which says that an energy change and its duration cannot be exactly measured simultaneously. The imaginary experiment features a light-tight box. The box contains a clock that is set to quickly open and close a shutter. Also contained is a gas of photons. Opening the shutter releases a single photon. By comparing the before and after weights of the box, one could determine the mass (and so the energy of the photon) with as much precision as desired. Einstein concluded that indeterministic quantum theory had to be rejected. After a sleepless night, Niels Bohr found a flaw in the hypothetical. When the photon escapes, it imparts an unknown momentum to the box. This kick causes the box to move in the gravitational field. By Einstein’s own theory of general relativity, movement in a gravitational field affects the rate of the clock. Since the clock is being affected by the unknown momentum, the time it measures inherits the uncertainty. So rather than refuting Heisenberg’s principle, the “clock-in-the-box” confirms it.

A popular tactic against counterfactuals is the method of counterdescription. In chapter 2, the method was illustrated with Bernard Williams’s redescriptions of “body transfer” cases. Another neat reversal is found in David
Cole's critique of the Chinese Room: Searle had deemed it plain that he would not understand Chinese even though he could transform strings of symbols expressing Chinese questions into strings expressing Chinese answers. Cole insists that Searle would understand Chinese. Cole first points out that the symbol manual and writing paper are red herrings; Searle could in principle memorize the rules and speak with the same fluency as a native. Cole's next point is that although bilinguals can usually translate, their failure to do so would not entail that they do not know the languages. They would only be abnormal bilinguals; linguistic schizophrenics who lack an ability most have. Compare them with aphasics who have peculiar combinations of abilities and disabilities. Thus, by redescribing the situation, Cole makes the attribution of understanding less outrageous and so can challenge Searle's assessment of the counterfactual.

Sometimes the counter-thought-experimenter only aims for a draw and can even target a counterfactual implied in an executed experiment. Both of these uses are illustrated by Mach's critique of the experiments, observations, and thought experiments deployed by Newton as evidence for absolute motion. Absolute motion is movement relative to absolute space. Relative motion is movement relative to a material object. Suspicion of absolute space led relationists to argue that all motion is relative motion. Newton argued that acceleration is absolute motion because no material object could serve as the frame of reference. (Both sides agreed that movement had to be relativized to something—they just disagreed over the relata. Saying 'x moved but did not move relative to something' was compared to 'x drank but did not drink something'.)

The centerpiece of Newton's case was a simple experiment that he actually performed (Figure 6.1). Newton hung a bucket by a cord and turned the bucket round and round until the cord was strongly twisted. The bucket was then filled with water. This completes the first stage of the experiment. Newton began the second stage by letting go. The bucket whirled. For a while, the bucket was in motion relative to the water. But gradually the initially smooth surface of the water became concave as friction with the sides imparted a centrifugal force to the water. When the speed of the water caught up to that of the bucket, it was at rest relative to the bucket, giving us stage 3. At stage 4, Newton suddenly stopped the bucket. So once again the water was in motion relative to the bucket. Yet the situation is not the same as stage 2; the water at 4 has a curved shape. Can we completely describe the water's motion by taking the bucket as our frame of reference? No, because the water is in motion relative to the bucket at stages 2 and 4, yet the water's behavior is not equivalent: at stage 2 it is flat, and at stage 4 it is curved. Can we, instead, relativize to the laboratory, the countryside, or the earth itself? These candidates are disqualified by the pervasiveness of centrifugal effects. We need a referent that works even for huge rotating objects such as orbiting planets and galaxies. No ordinary material thing meets this requirement. Hence, we are forced to relativize to an extraordinary thing, namely, space itself. Newton stressed that this basic frame of reference is not just an abstraction from the material things occupying it: "Absolute space, in its own nature, without relation to anything external, remains always similar and immovable." Newton knew that many would hope to wriggle out of commitment to this weird thing by saying that he must have overlooked an ordinary reference frame that has yet to occur to anyone. To foreclose this possibility, Newton deployed a thought experiment that stipulates all these hidden relata away.
Imagine two globes that are tied together with a cord and then twirled about their center of gravity (Figure 6.2). If all motion is relative, then there would be no difference between stationary observation of the rotating globes and rotating observation of stationary globes. But only in the first case would the cord become tense; "And thus we might find both the quity and the determination of this circular motion, even in an immense vacuum, where there is nothing external or sensible with which the globes could be compared." Newton further supplements the case for absolute motion by appealing to the flattened poles of large rotating bodies such as the planet Jupiter.

Absolute space is unobservable and radically different from all familiar objects. Hence, Mach urged perseverance: if we cannot relativize to the bucket, then we should try relativizing to some other body or group of bodies such as the fixed stars. According to Mach, the fixed stars qualify as an inertial frame as long as we are willing to postulate a universal force whose magnitude depends on the acceleration and mass of the interacting objects and which can operate well over enormous distances; for we can then say that gravity rotates relative to the stars activates this force: the faster the water spins, the more the huge mass of the stars distorts the shape of the water. Sounds like a funny force? Well, most forces other than contact forces strike us as peculiar: gravity, electricity, magnetism. Certainly, Mach's force is no stranger than Newton's alternative of a force exerted by space itself. If space is changeless, how would we reconcile spatial force with Newton's third law of action and reaction? Mach's force also satisfied the positivist requirement of being verifiable. The force might be detected with a modified bucket experiment featuring a bucket with sides that were several miles thick. For the difference in local mass might then be large enough to make the water curve a bit more than with a thin bucket.

But isn't Mach's force refuted by Newton's globe-twirling thought experiment? Mach defends his use of the fixed stars by preaching intellectual humility. What basis have we for predicting what would happen to the cord connecting the globes in an immense vacuum? Mach writes, "No one is competent to predicate things about absolute space and absolute motion: they are pure things of thought, pure mental constructs, that cannot be produced in experience." He rubs the point in with a rhetorical challenge. If Newton were right, we could hold the bucket steady in absolute space and rotate the heavens around the bucket without thereby deforming the water. Does the Newtonian really think he knows what would happen under such unprecedented circumstances?

Newton's appeal to the oblateness of Jupiter is equally empty. We only feel that the Jovian bulge confirms absolute space as long as we presuppose absolute space. Mach applies this point to a hypothetical variant of the Jupiter observation devised by Carl Neumann. Suppose that as a planet rotates the rest of universe abruptly vanishes. Isn't it obvious that the planet would continue to be flattened at the poles? Mach answers with the charge that Neumann has "made here too free a use of intellectual experiment, the fruitfulness and value of which cannot be denied. When experimenting in thought, it is permissible to modify unimportant circumstances in order to bring out new features in a given case; but it is not to be antecedently assumed that the universe is without influence on the phenomenon in question." What qualifies as an important feature of a thought experiment varies with what one is trying to demonstrate. The point holds for regular experiments as well. We are free to use a mercury thermometer to investigate freezing points but become ensnared in circular reasoning if we use it to demonstrate that objects expand with rising temperature. The theory describing the behavior of our experimental apparatus must be independent of the theory it is used to test.

4. Pseudo-absurdity. When the absurdity is a contradiction, it is not open to plausible challenge. But when the consequence is surprising without being contradictory, one has the option of denying that the consequence is an impossibility. The four ways of doing this parallel the four ways a defendant can plea against an accusation. I shall lead with the boldest and work down to the mildest.

The first submove is to deny even the appearance of absurdity. This is analogous to the claim that the court has no grounds for thinking a crime has been committed. In its boldest form, the denial of apparent absurdity amounts to the countercharge that the objects are insincere. Less feisty is the claim that there is no legitimate appearance of absurdity. Recall how Mach replied to Neumann's thought experiment by insisting that the consequence is strange only to those who beg the question. Alternatively, we may agree that the consequence is strange but only to one under the spell of a seductive misinterpretation.

A second option is to "bite the bullet"—agree that the consequence is strange but insist that it is a veridical rather than a falsidical paradox. This resembles the legal defense of justification. There are many famous examples of apparent absurdities that turn out to be surprising truths. For instance, the fact that the earth goes around the sun was long called the Copernican Paradox.

The analogue to the third tactic is to grant that a crime has been committed but to pin the blame on someone else (e.g., I may admit that the banjo in my attic is stolen but insist that I unwittingly "purchased" it from a thief). The defense as applied to thought experiment can be put under the slogan "Strangeness in, strangeness out". For the strangeness of the result is traced to the strangeness of the supposition. This strategy has been applied to a thought experiment inspired by an ancient tale of fatalism (a variant of which appears in Somerset Maugham's play Sheppy.) The story begins with a man who sees Death in Damascus. Death is startled but then warns "I am coming for you.
tomorrow." So the man takes the first camel out of town. The next day, Death finds his victim in Aleppo. "But I thought you'd be looking for me in Damascus!" sputters the exhausted traveler. "Not at all," replies Death; "that is why I was surprised to see you in Damascus—I knew I was to find you in Aleppo."

The thought experiment modifies this story by stipulating that the traveler knows that he has an appointment with Death. He also knows that the date is fixed weeks beforehand and that Death's appointment book is virtually infallible. Hence, although the traveler's choice does not cause Death to be at the appointed place, his choice is excellent evidence of Death's location. For the sake of simplicity, we further suppose that the traveler must choose between one of two places, Aleppo or Damascus. Notice that the traveler can foresee that he will regret whichever choice he makes. If he chooses to go to Aleppo, Death will meet him there, and the traveler will wish he had gone to Damascus. If he chooses to stay in Damascus, Death will meet him in Damascus, and he will wish he had gone to Aleppo. But since it is irrational to do what you know you will regret, the traveler is doomed to irrationality. This strikes most people as absurd, because they believe that if you rationally ought to do x, then you can do x. Allan Gibbard and William Harper parry the attack by arguing that we instead give 'ought' implies 'can'. They say the traveler is just a victim of freak circumstances that force him into an irrational choice. Hence, Gibbard and Harper contend that the absurdity has been fostered by the thought experimenter, not the decision theorist.

The analogue to the fourth defense against the charge of absurdity is to excuse the action on the grounds that it was the lesser evil. (This is popular among those responding to Shoemaker's fission case mentioned in chapter 1.) More specifically, this gambit appeals to the conservation of incredulity. Grant that the consequence of your position is hard to believe but emphasize that any position on the issue will have a counterintuitive consequence. Exhort your audience to concentrate on the relative counterintuitiveness of your solution rather than its absolute counterintuitiveness. There is a general rationale for this comparative theme. When faced with a paradox, one has a conflict of loyalties, therefore there must be a painful disownment of a previously cherished belief. You can redistribute the cognitive shock associated with the paradox, but you cannot eliminate it.

5. Impossibility theorem. According to Aristotle, terrestrial objects move only when pushed or pulled by an outside mover. The velocity of the object depends on the degree of push or pull and the resistance offered by the object and the medium. Hence, Aristotle's "equation of motion" was \( v = F/\mu \). Critics posed a hypothetical featuring an arrow traveling in a vacuum. If the equation held for this arrow, the arrow would either be motionless because the driving force was zero or it would fly at infinite velocity because the resistance to motion was zero. But since both of these outcomes are physically impossible, Aristotle's theory of motion is mistaken. In our regimented form, the paradox for Aristotle is as follows:

1. Aristotle's theory of motion is correct.
2. If so, then velocity must equal force divided by resistance.
3. If \( v = F/\mu \) and an arrow were in a vacuum, either \( v = 0 \) or \( v = \infty \).
4. Neither \( v = 0 \) nor \( v = \infty \) is possible.
5. It is possible to place an arrow in a vacuum.

Aristotelians responded that the content of the thought experiment was impossible; arrows cannot move in vacuums because empty space is impossible.

Since they were willing to accept the remaining four propositions in the quintet, the Aristotelians turned the thought experiment into a principle of sorts by regarding the four propositions as premises of an argument that concludes with the negation of the fifth as an impossibility theorem. A modern instance of this modal opportunism is George Schlesinger's handling of Newcomb's problem. He argues that if there were a Predictor of human decisions, rationality would require the one-box choice and rationality would require the two-box choice. But since this is absurd, it follows that the Predictor is impossible and consequently we are free agents.

This flexible use of the set of propositions constituting the thought experiment invites a spurious charge of radical inconstancy. Recall Charles Schmitt's criticism of the medieval debate over vacuums, discussed in chapter 2. The vacuists urged that a void could be created by stretching an impregnable container so that its volume was enlarged without the addition of new content. Thus, they were using the scenario to attack the principle that nature abhors a vacuum. Schmitt complains that the plenists used the same scenario to defend this principle. The aura of arbitrariness is reduced when we distinguish between fortifying your position and disarming your adversary. A fortifying defense supplies evidence that reinforces the principle. A disarming defense undermines an attack on the principle. (It is one thing to falsify a premise of your adversary's objection and another to falsify the conclusion of the objection.) The plenists were only using the thought experiment in a disarming defense of "Nature abhors a vacuum." They wanted to undermine the vacuist's premise that the tasks of stretching and sealing are cohesive. The plenists were willing to grant that you could build an impregnable container, and they were willing to grant that you could have a force powerful enough to stretch any container. What they denied was that you could have both simultaneously. Proving this would give the plenists an impossibility theorem that would remove an entire class of threats to their position.

Hidden impossibility results are often obtained by sweating more detail out of the supposition. This cross-examination technique is frequently used to debunk schemes for violating limits. For instance, one proposal for transcending the speed of light begins with a long rigid rod growing out into space from the equator. The longer the rod, the faster its tip whips about, because the earth is a rotating sphere. Therefore, if the rod continues to grow, the tip must eventually move faster than the speed of light. Physicists counter with close
questioning. Where does the material for the rod come from? If it is being drawn from the earth, the conservation of angular momentum requires that the angular velocity of the earth–rod mass decreases in the way a skater slows down as he extends his arms. As more of the earth turns into the rod, the tip slows; once all of the earth is converted into a celestial wand, the tip will be very slow. On the other hand, if there is an extraterrestrial source for the rod's material, then the additional mass must be given extra kinetic energy to keep it moving in pace with the angular velocity of the earth. But we will run out of energy before the tip reaches the speed of light. Hence, the appearance of possibility is a product of intellectual laziness.

Modal inquests are powerful but perilous stimulants. Close interrogation tends to inflate standards of possibility. Is it possible for an iron bar to float on water? That's a bar, not a needle, right? So we can rule out surface tension. Genuine H₂O? Do you mean real iron or an iron look-alike? Bear in mind that Fe has a specific gravity of between 7.3–7.8. And you meant float, didn't you? To float means for a thing to have a lower specific gravity than its medium. Since water has a specific gravity of 1.0, it only seems like an iron bar can float on water. Think it through, dummy!

Notice the slide from mere logical possibility to chemical possibility. The same equivocation of standards is evident in dialogues with determinists. You say that you could have chosen to read a different book? But didn't your choice issue from a particular set of beliefs and desires? Aren't those psychological properties a product of your genetic makeup plus your upbringing? Just when did your future become unfixed, open, free? The determinist traps his interlocutor by subtly changing the question. We begin with issue of whether alternate book selection is consistent with a loose set of laws and initial conditions and wind up with the question of whether the selection was consistent with a highly specific set of laws and initial conditions. By supersaturating the context with detail, one makes the event inevitable.

Sometimes the denial of possibility amounts to a charge of meaninglessness, rather than falsehood. This was a favorite among the logical positivists. Their verification criterion of meaningfulness declared that a statement is meaningful if and only if it is either analytic or verifiable. For example, 'All vixens are female' is meaningful because we can check the rule for 'vixen' in a dictionary and then deduce the truth of the statement. 'The earth bulges at the equator' is verifiable because we can confirm or disconfirm it with a geographical survey. Contrast these claims with 'Ghosts exist' as uttered by a believer who has de-empiricalized the notion of ghosts with the qualifications that ghosts are intangible, invisible, inaudible, and in general, undetectable. Given all these meaning-leeching provisos, we scornfully conclude there is no difference between there being ghosts and there not being ghosts. The claim only appears to have content. It is actually meaningless, just as the verification criterion dictates. The positivist goes on to scorn 'The universe doubled in size last night' and 'The universe popped into existence five minutes ago' as especially deceptive pseudostatements because they more closely mimic meaningful ones. At this point, the positivist draws our attention to the general pattern of sceptical thought experiments. From the assumptions that knowledge implies certainty and certainty implies the impossibility of error, the sceptic infers that knowledge is undermined by the mere possibility of mistake. Hence, the sceptic tries to refute the claim that you know you are holding a book by appealing to the possibility that you are dreaming or that your brain is being stimulated by a deceiving neurosurgeon or whatnot. These doubts are designed to be eliminable by empirical tests. So we seem mired in uncertainty and hence profound ignorance. But now the positivist advises us to turn this problem into the solution. Applying the verification criterion lets us eliminate these counterpossibilities as being meaningless, rather than false.

Critics of thought experiments have also harnessed a contrast principle; a term has application only if its opposite has application. The problem of evil provides the backdrop for an illustration. Some theists clamor for details about how the universe could be better than it is all things considered. David Hume responded to the challenge by envisioning a world in which there is no pain, only various levels of pleasure. Since putting your hand on a hot iron would cause your bliss to suddenly drop to mild euphoria, you would withdraw your hand to gain the greater pleasure. Thus, your body is protected from injury by differential pleasure, thereby eliminating the need for pain. Many theists charge that Hume's thought experiment is nonsensical because 'pleasure' only has meaning by virtue of its contrast with pain.

The appeal to meaninglessness lives on; witness Putnam's treatment of the sceptical worry that I am a brain in a vat. Putnam's causal theory of reference requires the speaker to have appropriate causal connections with his referent. But if I were a brain in a vat, I would lack this relation and therefore my utterance would fail to express the worry. Hence, if I am a brain in a vat, I cannot even think that I might be a brain in a vat!

Those with no love for verificationism can sometimes achieve a similar dissolution with an appeal to unknowability. Many thought experiments have a hidden epistemic component; that is, in addition to assuming the possibility of the predicament, they assume the possibility of the character's knowing his predicament. This assumption is challenged by one analysis of the Death case. It concedes that the traveler's predicament is possible but denies that the traveler can realize that he is in the predicament. The reasoning is that a chooser cannot pick an option that he knows to be inferior to the other alternative. So although outsiders can know that the traveler will meet Death either way, the traveler inevitably interprets the situation differently.

Since impossibility comes in different grades, there are varying grades of unknowability. This opens the possibility of mixed solutions to paradoxes. Consider a hypothetical that arose from the policy of mutual assured destruction ("If they nuke us, we'll nuke them"). Suppose that the only way to deter an evil is to threaten a fiendish retaliation. This threat of a counterstrike must be sincere because your adversary detects bluffing. Is it moral to form the conditional intention to do the immoral deed? The question arouses mixed feelings; for the intender seems virtuous in that his intention prevents evil yet vicious in that he is prepared to do evil. So which is it? David Lewis holds that there is no
determinate answer to this speculative question. Our moral code could be precisified to come down on either side of the issue. The shock caused by either answer could be explained away as a pseudoabsurdity generated by the peculiarity of the scenario. This may seem an evasion. Doesn’t the existence of the current U.S. policy on nuclear warfare show that the scenario is realistic? Granted, our information is more limited in the real world. But can these uncertainties be set aside with minor idealizations? Lewis answers that our ignorance is intractable. There are surprisingly deep reasons why the decision-makers could not know enough about how much damage had been sustained during a first strike or what the effect of a retaliatory strike would be. So Lewis adopts an impossibility solution for “realistic” variants of the hypothetical and a pseudoabsurdity solution for the more far out possibility where the decisionmaker does have the knowledge.

C. Summary of Necessity Refuters

I close with a thought experiment that illustrates four of the five resolutions of a necessity refuter thought experiment: the St. Petersburg paradox. The mathematical theory of rational choice developed alongside advances in probability. At the time it seemed that if rational choice were amenable to mathematical treatment, rational agents must maximize their expected value. For instance, if you are offered a prize of $10 on the condition that a coin lands heads, then the offer is worth $5, because you have a probability of 1/2 of gaining $10 and 1/2 × $10 = $5. An offer of a $24-dollar prize for a die yielding a six has a lower expected value because 1/6 × $24 = $4. Hence, the wise chooser takes the coin offer rather than the die offer because he should maximize expected value. But now suppose you receive a more complex offer. A fair coin will be tossed until one head results. You will then be paid $2^n where $n$ equals the number of tosses. Hence, the expected return is (1/2 × $1) + (1/4 × $2) + (1/8 × $4) + . . . + (1/2^n × $2^n) . . . Since each addend equals 50c and there are infinitely many of them, the sum is infinite. Thus, someone who maximized expected money should be willing to pay any amount of money for this bet. Yet common sense tells us that the offer is not worth much at all. Few people would pay $1,000 for the deal. So the belief that rational choice is amenable to mathematical analysis seems to lead to an absurdity.

The St. Petersburg paradox led to a rich set of reactions. A few accepted it as a cogent refutation of the idea that free choice has a mathematical structure. Some mathematicians defiantly declared that a prudent man would risk everything on the bet. They claimed that the result was a pseudoabsurdity; it only looks absurd because we are only familiar with bets offering finite gains.

Condorcet and Poisson rejected the thought experiment on the grounds of impossibility. If heads does not come up by the hundredth toss, the gambler should receive a profit equivalent to a mass of gold bigger than the sun. So the deal cannot be kept. Others extended this criticism by insisting that money is essentially finite because money has exchange power only to the extent that it is scarce.

Daniel Bernoulli challenged the counterfactual ‘if an expected utility maximizer were offered, then he would be willing to pay any finite price to participate’; that is, Bernoulli denied that the principle of maximizing expected utility implied that the deal was of infinite value (even if infinite money were possible). He pointed out that doubling one’s cash holdings from a million to two million does not really double the value to you. Each new dollar tends to have less influence on your welfare than the preceding dollar. Bernoulli’s insight is enshrined in contemporary economics as the law of the diminishing marginal utility of money. The rate of diminution resists precise calculation, but Bernoulli persuasively argued that it was a logarithmic function which prevents the sum from being infinite.

D. Possibility Refuters

A minority of thought experiments vary from the format introduced at the beginning of the chapter. Instead of connecting the modal source statement to a necessity, they connect it to a possibility statement. And instead of affirming the possibility of the thought experiment’s content, they say the possibility statement implies its copossibility with the content. The two revisions yields a new quintet of jointly inconsistent propositions:

1. S
2. S ⊡ O1 Possibility extractor. This proposition draws a possibility consequence from the source statement.
3. (I & C) ⊡ W
4. ~O W
5. O1 ⊡ O(I & C) Content copossibility. This asserts that the statement extracted at ii is true only if it is compatible with the content of the thought experiment.

Three of the five members are unchanged from the necessity refuter schema. They do not need further explanation. But the change at steps ii and v betokens an important shift in tone. Necessity refuters say that the source is too closed minded, that it rules out genuine possibilities. Possibility refuters say that the source is too open-minded, that it saddles us with spurious possibilities.

1. Bad source statement. We can show how possibility refuters are supposed to work with the help of a standard objection to polytheism.

1. Polytheism: more than one god exists.
2. If polytheism is correct, then two omnipotent beings can coexist (because gods are perfect and perfection implies unlimited power).
3. However, if there were two omnipotent beings and they were to have a shaving match, an irresistible force would meet an immovable
object. (For one would have the power to move anything and the other would have the power to resist any movement.)

iv But it is impossible for an irresistible force to meet an immovable object: either can exist, but it is contradictory to say they coexist.

v If it is possible for two omnipotent beings to exist, then it is possible for them to have a shoving match.

Something has to go! Most theologians dump polytheism, thereby agreeing that the number of gods is either 1 or 0. Thus, the major battle is between monotheism and atheism.

Purveyors of possibility refuters often engage in overkill; they make an already absurd consequence more absurd. Mach is responsible for a charming example of this magnification. His thought experiment demonstrates the physical transitivity of the equal-mass relationship, that is, that two masses that are equal to a third mass are equal to each other. Suppose that A, B, and C are elastic bodies that freely move along a smooth, rigid ring (Figure 6.3). Further suppose that the principle fails to hold for these three objects. In particular, assume that A has a mass value equal to B upon collision and that B has a mass value equal to C but that C has a greater mass value than A. If we then impart a velocity to A, A will pass the velocity to B, and then B to C. But since C has a greater mass value than A, its collision will give A more velocity than A initially had. This surplus will then be carried to B, then to C, whereupon another boost will be given to A. So on each cycle, the bodies go faster and faster without limit. This endless amplification violates the first law of thermodynamics, which says that energy is neither created nor destroyed. So Mach concludes that ‘equal mass’ must be transitive. But notice that even one cycle around the ring is precluded by the conservation of energy.

Economists use the money pump argument with the same zeal to prove the transitivity of indifference: if a rational agent is indifferent between A and B and indifferent between B and C, then he is also indifferent between A and C. Suppose indifference were not transitive. It would then be possible for a rational agent to be indifferent between A and B and indifferent between B and C yet prefer A to C. But if this were so and if the agent, Sap, were offered the following trades, he would be turned into a money pump.

<table>
<thead>
<tr>
<th>Sap</th>
<th>Trader</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. Initial holdings:</td>
<td>B, C, $100</td>
</tr>
<tr>
<td>1. Sap trades C and $3 for A</td>
<td>A, B, $97</td>
</tr>
<tr>
<td>2. Sap trades B for C and $1</td>
<td>A, C, $98</td>
</tr>
<tr>
<td>3. Sap trades A for B and $1</td>
<td>B, C, $99</td>
</tr>
</tbody>
</table>

Notice that Sap has traded for a net loss. Continued trading in this pattern eventually results in all of Sap’s money being lost to the merchant. The agent will have been turned into a “money pump.” But no rational agent can be turned into a money pump. Yet if it is possible for a rational agent to have intransitive indifference, it is possible for this preference structure to occur in the trading context. Hence, rational indifference is transitive. Ducky, but once again notice that a single cycle suffices to demonstrate Sap’s irrationality.

Since ‘redundant’ is relative to belief systems, there are thought experiments that legitimately amplify absurdity. One targets microphysicists who maintain that at the quantum level, things are in an indeterminate state until the observation is made. Erwin Schrodinger ridiculed this idealist stance by imagining a cat that is sealed in a box containing a capsule of poisonous gas. The gas will be released just if an emission of a radioactive source triggers it within a certain interval. After the period elapses, we wonder whether the cat survived, so we believe ‘Either the cat is alive or dead’. But given that the particle from the radioactive source only acquires definite properties upon being observed, it follows that the cat itself is in an indeterminate state, neither dead nor alive! Looking inside the box settles its fate, but it has no fate until we peek. But obviously, the only indeterminacy here is epistemological. We do not know whether the cat is dead or alive, but it is surely one or the other. Many microphysicists eat this absurdity. So Eugene Wigner hardened the bullet by substituting a conscious human observer for the cat. “Wigner’s friend” may be unneeded for the persuasion of some physicists but necessary for the persuasion of those with a different background or temperament. Although our sense of absurdity has the biological uniformity suggested by Mach, there are also systematic variations. Thus, effective thought experimenters throw a wide net.

2. Misconnection. Archytas of Tarentum attacked Aristotle’s doctrine of a finite universe with a thought experiment that influenced cosmologists for two thousand years. If the universe is finite, then it is possible for a man to travel to its edge. If he were then to throw a spear, it would rebound or vanish. But that’s crazy. So the universe must be infinite. Einstein attacked the second step of the thought experiment: finitude does not imply a boundary, because space could be curved.

For a contemporary illustration, consider the reasoning behind those who say the existence of a universal speed limit should not shock us because the
opposite possibility is at least as absurd. If there were no maximum speed, an object could travel at infinite speed. If this infinitely fast object made a grand tour of every spot in the universe, it would be everywhere simultaneously. But an omnipresent object is absurd. The trouble with this thought experiment is that the absence of a maximum does not imply an infinite. (For example, there is no largest fraction of the form \( n/(n+1) \) even though none of them is infinitely large.)

3. *Erroneous counterfactual.* One common objection to time travel is that time travelers could change the past; for suppose that Tina the Time Traveler goes back to the day when her father was a toddler. Now it seems that Tina could grab his throat and strangle him. But if he dies, Tina would not get born and so could not be around to do the dirty deed. Thus, the objection is that if time travel were possible, the time traveler could depart for the past in 1992 and then alter what happened in 1942. But since what has already happened cannot be changed, time travel is impossible.

Time travel buffs reject the counterfactual. They agree that the past cannot be changed but deny that this would make the time traveler *miraculously* impotent; for if time travel is possible, then Tina the Time Traveler is already part of the past; her contribution is preembedded in the course of history. Hence, when Tina's fingers curl about the throat of her toddler father, something will prevent the patricide: he runs away, Tina's grandfather enters, she has a change of heart, or some such. We do not know exactly what prevents it. But we know that something must prevent it, because we know the toddler lives to sire Tina the Time Traveler.

4. *Pseudo-absurdity.* The fourth move is to deny that the apparently absurd consequence is really absurd. As with the necessity refuter schema, there are four submoves corresponding to the four ways of defending yourself against a charge of wrong doing.

First one can deny the appearance of absurdity. Consider Mark Johnston's objection to Robert Nozick's closest-continuer theory of personal identity. According to this theory, the closest among a field of sufficiently close continuers of an individual is that individual. Johnston objects that this theory would make one's identity dependent on future events. Suppose that a machine has transferred the psychology of a person who began with body A into body B. The person in the B-body gets up, walks about, and says to himself "I am A. I did not just come into existence." He then notes that the machine is reading the A-psychology into body C and will soon make the person with the C-body a better continuer than B. So he turns the machine off, thereby making himself the best continuer. According to the closest-continuer theory, he has thereby made his past thought "I am A" true. But Johnston stresses that the relevant facts were settled before the B-body person could decide whether to turn off the machine. Some defenders of the closest-continuer theory respond that this consequence only appears absurd as long as we overlook the fact that the future determines the past just as much as the past determines the future.
clock sent off in a speeding spaceship lags behind an earthbound clock. Many people found this incredible. The absurdity was given a human face by substituting twins for the clocks. If acceleration was sustained long enough, the speedy twin could return to find his brother an old man. Relativity theorists treated the absurdity as a veridical paradox. They have since gained empirical support from the fact that atomic clocks sent on speeding planes lag slightly behind ones left on ground, from the increased lifetimes of highly accelerated mesons, and from slight inaccuracies in Doppler’s original formula for the wavelengths of light emitted by rapidly moving atoms.

Richard Dedekind’s definition of ‘infinite set’ embodies a pseudoabsurdity response to a paradox discussed by Galileo. When we say that the number of As equals the number of Bs, we mean that each A can be uniquely paired with each B—that there is a one-to-one correspondence. However, cases involving infinite sets suggest that this account of equality is too broad; for if it were true, then an infinite subset of a set would be as big as the set itself. As Galileo noted, the natural numbers can be set off in a one-to-one correspondence with those that are squares:

Natural numbers 0, 1, 2, 3 \ldots n \ldots

Squares 0, 1, 4, 9 \ldots n^2 \ldots

However, the squares are included in the natural numbers. Moreover, the squares get rarer and rarer as one progresses down the number line. In any case, it’s absurd for a set to have a proper subset whose members can be put in a one-to-one correspondence with the set itself. However, Dedekind accepted the “absurdity” and even enshrined this paradox as the *definition* of ‘infinite set’.

Third, one can reply to a charge of absurdity by saying that the absurdity is inherited from the thought experiment’s weird scenario. This is how hedonistic utilitarians tame utility monsters. Suppose that an individual was able to convert resources into pleasure much more efficiently than anyone else, at a nondecreasing rate, and without limit. Would we then be obliged to turn over all our assets to this lean, mean pleasure machine? Yes, says the hedonistic utilitarian, but the absurdity can be completely traced to the weirdness of the utility monster. If you ask a funny question, expect a funny answer.

5. *Composibility* denial. One class of thought experiments misinfers a possibility statement by means of the following any/all fallacy:

1. Any F is possibly a G. \( \forall x \exists y \forall F x \)
2. It is possible for all Fs to be Gs. \( \forall x \exists y \forall F x \)

The invalidity of this argument form is demonstrated by its invalid instances. Any New Jersey lottery ticket is possibly a winner, but it is impossible for all of the tickets to be winners. Now consider a thought experiment designed to show that personal identity does not require consciousness. We first assert that no thought is essentially a conscious thought. Then we reason that since any thought could be unconscious, all of one’s thoughts could be unconscious. But yet you would still exist. The suspicious step here is the any-to-all inference. Perhaps an unconscious thought is just one that has a low degree of accessibility as compared to the average accessibility of one’s thoughts. Then although each thought could be unconscious, it would be impossible for all of them to be unconscious—they cannot all be below average!

A second example concerns a thought experiment designed by Michael Scriven to refute predictive determinism. Predictive determinism says that every event can be predicted beforehand. Scriven has us suppose that there is an agent whose dominant motivation is to avoid prediction (Avoider). Should this countersuggestible fellow learn that Predictor says he will drive to town, he will stay home. Therefore, to predict what Avoider will do, Predictor must conceal his prediction. Unfortunately, Avoider has enough data, laws, and calculating capacity to duplicate Predictor’s reasoning and thus his prediction. Hence, Avoider is unpredictable and therefore a counterexample to predictive determinism.

David Lewis and Jane Richardson object that Scriven has falsely assumed that Predictor and Avoider can simultaneously have all the needed data, laws, and calculating capacity:

The amount of calculation required to let the predictor finish his prediction depends on the amount of calculation done by the avoider, and the amount required to let the avoider finish duplicating the predictor’s calculation depends on the amount done by the predictor. Scriven takes for granted that the requirement-functions are *compatible*: i.e., that there is some pair of amounts of calculation available to the predictor and the avoider such that each has enough to finish, given the amount the other has.

Lewis and Richardson maintain that this compatibility assumption is only tempting as long as we are under the spell of the ambiguity of ‘Both Predictor and Avoider have enough time to finish their calculations’. Parsing the sentence one way yields a truth—that against any given avoider, the predictor can finish and that against any given predictor, the avoider can finish. But the compatibility premise claims that they can finish against each other.

E. Summary of Possibility Refuters

The paradox of the stone illustrates four resolutions of the possibility refuter thought experiment. If God exists, then it is possible for an omnipotent being to exist. If there were an omnipotent being and he tried to make a stone so big that he himself could not lift it, then he would succeed (because he can do anything) and not succeed (because success implies a contradiction). But no one can both succeed and not succeed at a task. Yet if an omnipotent being is possible, he can try to make unliftable stones (because omnipotence ensures he can try to do anything).
Atheists are free to reject the source statement and conclude that God does not exist. A few theists (such as the Boston personalists) challenge the modal extractor by denying that God is omnipotent. Descartes is the most famous of a small band of thinkers who take omnipotence to mean that God’s power is not even limited by the laws of logic. Members of this group might therefore deny that the contradictory feat is absurd.

However, the most popular theological response is to deny the counterfactual. In particular, theologians reject the following element of the larger counterfactual: 'If an omnipotent being were to try to make a stone too large for him to lift, then he would succeed.' Their case proceeds by first getting the promitor of the paradox to agree that contrary to Descartes, omnipotence does not imply the ability to bring about contradictory states of affairs. Otherwise, there is a much simpler objection to omnipotence: no one can make a round square. Omnipotence has to be understood as the ability to do whatever is logically possible. Hence, God is omnipotent if he can bring about any consistent state of affairs. We can have unliftable stones because the stone’s weight only has to exceed our finite lifting ability. So ‘unliftable stone’ is a consistent term relative to us. However, it is not consistent when relativized to an omnipotent being. Our anthropocentrism misleads us into thinking we have specified a consistent task for God. Hence, the answer to ‘Can God make a stone so big that He Himself cannot lift it?’ is ‘No, but so what?’ Omnipotence does not imply that He could lift it.

III. The Identity Conditions for Thought Experiments

Objection: My analysis underestimates the flexibility of thought experiments. For many thought experiments are modified to make new points and to meet objections. These changes can only be formally reflected by replacing a proposition in the set constituting the thought experiment. But since a set is defined in terms of its members, you no longer have the same set. Therefore, we are forced to say that any alteration of a thought experiment yields a new thought experiment. However, we can easily demonstrate that looser standards prevail.

For instance, Bertrand Russell’s five-minute hypothesis has been mounted by different thinkers to make different points. Russell introduced it in 1921 to refute ‘If there are memory-beliefs, then there were past events’. By supposing that the universe popped into existence five minutes ago complete with fossils, ‘memories’, and so forth, Russell showed how it is possible for memory-beliefs to exist without a past. This boosted scepticism about our knowledge of the past. One might think the universe could be shown to be older than five minutes by applying standard dating techniques. Cut down a tree. Count ten rings. Conclude that the universe is at least ten-years-old because a new ring forms once a year. But this begs the question. The five-minute hypothesis grants that things look like they have been around for many years. Dating methods presuppose, rather than prove, a past. Hence, Russell’s thought experiment propels an interesting sceptical issue. But others retreat the five-minute hypothesis for purposes not envisioned by its inventor. Peter Unger has used Russell’s brainchild to attack the causal theory of reference. The causal theory says that names refer to their bearers by virtue of appropriate causal relations leading from the event of naming the bearer to the speaker using that name. Hence, if the universe popped into existence five minutes ago, names only appear to refer to their bearers. Since Unger finds this consequence absurd, he rejects the causal theory of reference. Causal theories of knowledge, meaning, and intention have attracted parallel criticisms. The five-minute hypothesis has also been deployed against etiological accounts of biological function. These theories say that an organ’s function is the effect for which it was selected. For instance, the function of a human heart is to pump blood because human beings survive because of that effect. But if the universe popped into existence five minutes ago, organs lack histories of selection. Etiological accounts therefore imply that our organs would then lack functions! Russell’s five-minute hypothesis is also used in ethics to contrast “forward-looking” utilitarian theories of justice with “backward-looking” retributivist theories. The utilitarians say that an act is right exactly when it maximizes goodness. Since we can only affect the future, this amounts to saying that right actions maximize good consequences: the past is irrelevant. Therefore, the truth of the five-minute hypothesis would be morally irrelevant. Deontological ethicists hold retributivist theories of morality saying that an injustice is done whenever someone is punished for a crime they did not commit. If the universe popped into existence five minutes ago, then so did the inmates of Attica. The prisoners would be the (albeit unwitting) victims of injustice because their punishment is undeserved.

The only proposition that remains constant through this variety of applications is the C part (the content saying that the world just popped into existence). Yet we are not misspeaking when we say that Russell’s thought experiment has been reused for different purposes. In ordinary usage, content continuity appears to be a sufficient condition for identity: same scenario, same thought experiment. However, content continuity is not a necessary condition. In the sixth century, Simplicius quotes Archytas’ spear-throwing objection to finite space and asks what would happen if the man at the edge of the universe stuck out his hand. Since there would be no resistance, the man could take another step and stick his hand out again. Since the peripatetic frontiersman could continue this hand extending process without limit, the universe is infinite. Simplicius’ scenario is a bit different from Archytas’, but many commentators insist that it is the same old thought experiment.

We freely revise the content of thought experiments to overcome sticking points and count ourselves as amending rather than replacing the original thought experiment. Witness how news that a scenario is logically impossible is rarely treated as lethal. The normal inference is that the thought experiment has developed a glitch, so that we need to tinker to bring it back on-line. Thought experiments that run afoot of recent science readily adapt to the new climate of opinion. So instead of being delicate, set-theoretic butterflies, thought experiments are as tough as subway rats.
Often, the original thought experimenter is not in the best position to do
the repair work. An outsider may take over because he has the training, insight,
or time to make modifications. As other thinkers join in, the thought experi-
ment becomes a collective enterprise. Peter Galison has documented how the
high price of apparatus drives high-energy physicists toward collective experi-
ments. Apparatus-free thought experiments show how collectivism can
arise just through the need for intellectual division of labor.

Frequently, the thought experimenter will even revise the content to meet
objections he foresaw. This occurs when the thinker uses the dialectical process
as a vehicle for exposition. Indeed, the typical thought experiment undergoes a
combination of "canned" revisions and external modifications. Consider An-
thony Quinton's Two-Space myth. Its target is the metaphysical principle that
there can only be one space. If this principle is true, then it is necessarily the
case that every place can be reached from every other place by traveling
through intermediate places. Our confidence in this necessity underlies the joke
about the confused guide who, after several false starts, concludes, "You can't
get there from here!" Quinton challenges this strand of common sense with the
following scenario:

Suppose that on going to bed at home and falling asleep you found yourself to
all appearances waking up in a hut raised on poles at the edge of a lake. A dusky
woman, whom you realise to be your wife, tells you to go out and catch some
fish. The dream continues with the apparent length of an ordinary human day,
replete with an appropriate and causally coherent variety of tropical incident.
At last you find yourself awake at home to the world of normal responsi-
bilities and expectations. The next night by the side of the tropical lake
continues in a coherent and natural way from the point at which it left off. And
so it goes on. Injuries given in England leave scars in England, insults given at
the lakeside complicate lakeside personal relations... Now if this whole state
of affairs came about it would not be very unreasonable to say that we lived in
two worlds.30

The key counterfactual in our regimented version of Quinton's case is that if all
places are mutually accessible and you were to have these experiences, then
you would not be in a position to describe yourself as living in spatially isolated
worlds. You would be instead obliged to discount at least one set of your
experiences as delusory (but which?) or would have to persist in believing that
there is some hidden route from England to the lakeside (despite your most
exhaustive geographical investigations). Since Quinton thinks this epistemic
fate absurd but the sequence of experiences possible, he concludes that space is
not essentially singular.

So far, so good. Quinton's thought experiment snugly fits the necessity
refuter schema. The trouble is to accommodate its dialectical development. An
early objection to the Two-Space myth was that no one would believe your
report of a double life. Your audience's incredulity would give you ample
reason to discount the veridicality of at least one line of experience. Quinton
therefore amended the scenario so that most of the members of your commu-

nity report similar experiences. Indeed, their experiences correlate: each person
in England has exactly one counterpart in the lake district. (Perhaps they
identify each other with self-portraits or by making secret appointments to meet
in the other world.) A second objection was that the two-space explanation of
your experiences requires locating your body in two spots at a single time,
namely, a British bed and tropical boat. To avoid saddling your explanation
with the burden of this heresy, Richard Swinburne adds the friendly amend-
ment that folks who report double lives vanish from their beds soon after falling
asleep and reappear just before waking.31 He also helps by weakening the rival
explanation of there being a hidden route between the worlds by further
stipulating scientific differences between the world. For example, in England
gravity follows an inverse square law while in the lake district it follows an
inverse cube law. Swinburne's kibbitzing continues for a full chapter.

Evidently, the thought experiment can survive the revisions even if every
component and subcomponent of the original set of propositions is replaced. It
owes its survival to the resemblance between the new components and the old
as well as to its historical connection with the old thought experiment. Learn-
ing that William Newcomb invented Newcomb's problem after hearing the
prisoner's dilemma buttresses the case for their identity. The identity thesis is
further supported by David Lewis's logical demonstration that the prisoner's
dilemma is two Newcomb's problems "put side to side."32

The question 'When is a thought experiment the same thought experiment?'
has many of the complexities of other issues of identity: When is a ship the
same ship? When is a person the same person? Sometimes, as in patent law, the
question has practical significance; the significance in our case is theoretical.
Regardless of the niceties involved in specifying content continuity, proposi-
tional resemblance, and historical connection, we can see that the customary
standards of thought experiment identity are much more lenient than the set-
theoretic conception I have proposed.

The force of this objection is dissipated by pointing out that the same
difficulties arise for the logician's definition of 'argument': an argument is a set
of propositions of which one, the conclusion, is supported by the rest. Com-
mentators on the design argument, the private-language argument, and the
paradigm case argument do not treat any component as essential to the identity
of the argument. History and resemblance are as relevant here as they are to
thought experiment. Indeed, virtually all set-theoretic definitions are em-
barassed by the same flexibility.

The definitions are saved by the relativity of individuation: what counts as
the same F varies with one's purpose. The "conflict" between our ordinary
usage of 'same thought experiment' and our regimented usage is only an
illusion induced by equivocation. Relative to the purpose of establishing au-
thorship and priority, historical criteria dominate. Relative to the purposes of
theoretical unification and heuristics, propositional resemblance is more signif-
ificant. Since the purpose at hand is theoretical classification, maximally fine-
grained criteria are appropriate. Hence, instead of allowing a thought experi-
ment to survive a slight amendment, we say that there are two thought
experiments, one being a slight variation of the other. Two factors reinforce this strictness. First, unlike concrete objects, the abstract original can exist simultaneously with its altered version. Although a car owner cannot park his preprinted and postprinted automobile side by side for comparison, a poet can compare his original sonnet with its modified descendent. This difference makes us more apt to treat the modified abstract object as nonidentical to the original. Another factor favoring the set-theoretic definition of thought experiment is that ordinary speakers gravitate toward this fine-grained standard when pressed by the challenge ‘Is this thought experiment really identical to that one?’. This tendency toward pickier standards holds for all absolute terms (flat, free, certain), as noted in chapter 4. Since speakers usually overlook their slide to different standards, they often have the impression that the finest-grained standard is the uniquely correct one. This privileging illusion accounts for some of the persuasiveness of skepticism and hard determinism. My classification scheme will enjoy the same illicit boost! But honesty is the best policy. Therefore, I abjure this shadowy force and reiterate my claim to have found only a scheme suited to theoretical purposes, not the all-purpose truth of the matter.

IV. An Extension to Ordinary Experiments

Since most thought experiments could be turned into regular experiments by actually bringing about the envisioned scenario (the C part), we should expect a subset of normal experiments to be analyzable as paradoxes. Let us illustrate with Blaise Pascal’s demonstration that air is something, rather than nothing. He weighed an empty balloon, inflated it, measured a small weight gain, and concluded that air was a substance. We regiment it thus:

1. Source Statement: Air is nothing.
2. Modal Extractor: If air is nothing, then it must be weightless.
3. Counterfactual: If air is weightless and only air were added to a container, then the container would not grow heavier.
4. Absurdity: It is impossible for the inflated container to have not gained weight.
5. Content possibility: It is possible that only air was added to the container.

As before, the necessity refuter scheme lines up the possible challenges and indirect usages of the experiment. For example, one might reject (5) and take the experiment as evidence that an invisible impurity rushed in with the air.

Some critics might complain that the necessity refuter formulation dilutes the force of the experiment. Instead of (3), we should insert the stronger statement from which it was derived: only air was added to the container. However, the experimentalist should use the least controversial assumption necessary for the result. The actual insertion of pure air should be cited only as a lemma.

Another critic might object that proposition 4 overburdens the experiment: Pascal only tried to show that a particular balloon did in fact weigh more after inflation, not that it had to weigh more. My reply is that this criticism is also confusing salient intermediates with the key players. Scientists have little professional interest in facts that are only accidentally true. Even their simplest enumerative inductions are intended to establish the necessity of the concluding generalization:

1. Animal 1 is a platypus and has a bill.
2. Animal 2 is a platypus and has a bill.
   . . . . . . . . . . . .
100. Animal 100 is a platypus and has a bill.
101. All platypuses have bills.

If the conclusion is read merely as a universal generalization (not as a law), then it will not support the counterfactual ‘If animal 101 were a platypus, it would have a bill’. Without such counterfactuals, scientists have no hope of explaining why things are the way they are, rather than some other way. Pascal’s experiment fails to refute the source statement if the balloon merely happened to weigh more after inflation. It is only effective if understood as showing that the inflation made the balloon heavier—that the balloon had to get heavier. As Leonardo remarked, “Necessity is the guardian of nature.”

This logical similarity between a thought experiment and its executed counterpart is compatible with the existence of epistemological differences. Executing an experiment gives you better evidence for the fourth and fifth components of the necessity refuter scheme. As stressed in the chapter on armchair inquiry, both experiments and thought experiments are designed to make you an authority, but they try to do it in different ways. An experiment makes you an authority by enhancing your perceptual opportunities. An effective thought experiment amplifies your nonperceptual resources—such as recall; the transformation of one kind of knowledge into another; internal redelegation of cognitive tasks; and the elimination of cognitive obstacles, hindrances, and hang-ups. Since a regular experiment can also employ these nonperceptual strategies along with its perceptual strategies, one might expect it to dominate thought experiment. But this overlooks the advantage of specialization. Powered planes and gliders can both glide, but the different design emphasis leads us to expect gliders to glide better.

V. The Big Picture

My goal in this chapter was to present a taxonomy of thought experiment. By treating a thought experiment as a stylized paradox, we mature the idea that it reveals inconsistencies. We also expose the structure of our ensuing ambivalence, as well as the structure of resolutions. These benefits are achieved at the
price of artificiality. Often, we must reshape thought experiments to make
them fit the mold. Since this regimentation perturbs the customary exposition
of many thought experiments, the result will strike us as unnatural and even
perverse. However, the logical merit will be preserved even if there is some
aesthetic and rhetorical disfigurement.

My emphasis in this chapter has been on “global” aspects of thought
experiments. I craved a general framework that could subsume the particulars
in a systematic fashion. But I am also eager to detail the classification scheme
so that it can guide a fine-grained analysis of special kinds of thought experi-
ments. In particular, I wish to apply the framework in a way that will do full
justice to the sort of cases that intrigued Kuhn. Hence, chapter 7 opens up the
mechanism by which Kuhnian cases materialize on the intellectual landscape.

7

Conflict Vagueness and
Precisification

Creative workers need drink at night, ‘Roses and dung’. (Or: mathematicians
read ‘rubbish!’) An experimentalist, having spent the day looking for the leak,
had had a perfect mental rest by dinner time, and overflows with mental
activity

J. E. Littlewood

Now I want to pinpoint the property that excited Kuhn’s interest in thought
experiments. This property, conflict vagueness, often generates inconsistent
beliefs but is not itself inconsistency. Although it is absent from most thought
experiments, a substantial portion of the most provocative thought experi-
ments do spring upon this species of vagueness: for they motivate conceptual
reform by touching a nerve of indeterminacy. Hence, study of conflict vague-
ness reveals the ways thought experiments restructure our conceptual scheme.

I. General Features of Vagueness

A word is vague if and only if it has at least one borderline case. For example,
‘bald’ is vague because there are men who are neither clearly bald nor clearly
nonbald. The uncertainty created by borderline cases is irresolvable. No
amount of observation, experiment, or conceptual analysis can answer the
question. Borderline cases are inquiry-resistant.

If a term has an actual borderline case, it is extensionally vague. ‘Food’ is
extensionally vague because coffee is an actual borderline case of food. If a
term has a possible borderline case, then it is intensionally vague. ‘Mermaid’ is
intensionally vague because it is possible for a part woman–part fish to be too
fishy to qualify clearly as a mermaid yet too womanish to qualify clearly as a
nonmermaid. Since whatever is actual is possible, any extensionally vague term
is also intensionally vague—but not vice versa: that is the message behind
‘mermaid’. The extension of a term is the set of things to which it applies. The
extension of ‘vowel’ is {a, e, i, o, u} and that of ‘human’ contains you, me,
Manuel Noriega, and the rest of us.