What Are We to Think about Thought Experiments?

LAWRENCE SOUDER

Drexel University
Culture and Communication
3141 Chestnut Street
Philadelphia, PA 19104, U.S.A.
E-mail: ls39@drexel.edu

ABSTRACT: Arguments from thought experiment ask the reader to imagine some hypothetical, sometimes exotic, often fantastic, scenario for the sake of illustrating or countering some claim. Variously characterized as mental experimentation, imaginary cases, and even crazy cases, thought experiments figure into both scientific and philosophical arguments. They are often criticized for their fictive nature and for their lack of grounding. Nevertheless, they are common especially in arguments in ethics and philosophy of mind. Moreover, many thought experiments have spawned variations that attempt to both affirm and refute their original arguments. These emended thought experiments exhibit a variety of styles, details, and embellishments. A rhetorical analysis of these variations suggests a reciprocal influence between the arguers' selection of details and their philosophical commitments. I offer examples of this relationship from the variations on John Searle's Chinese Room thought experiment and Judith Thomson's unconscious violinist thought experiment.

KEY WORDS: Analogy, argument, Chinese Room, gendankenexperiment, hypothetical, narrative, philosophy of mind, rhetoric, thought experiment

> The subjects of our deliberation are such as seem to present us with alternative possibilities: about things that could not have been, and cannot now or in the future be, other than they are, nobody who takes them to be of this nature wastes his time in deliberation (Rhetoric, 1357a 5).

INTRODUCTION

Since Aristotle, legions of arguers have pressed for 'just the facts, Ma'am.' Yet, it seems that when we argue, we often do so over things that didn't happen and even things that couldn't. Hypothetical scenarios can so easily creep into our conversations, particularly when no real example to support our claim comes readily to mind, or when no real example is possible. Philosophers especially seem to find the use of counterfactual examples irresistible. In arguing over abortion, for example, R. M. Hare remarked, 'We might think differently of abortion if wombs were made of glass' (p. 214). Philosophers, however, haven't cornered the market on counterfactuals; the American Cancer Society's antismoking poster asks, 'If what happened on your inside happened on your outside, would you still smoke?' Indeed, Aristotle notwithstanding, even in our daily lives and in the privacy

Argumentation **17**: 203–217, 2003.

of our own heads, we often entertain what might have been if things were otherwise.

In fact, hypotheticals allow us to see the world through someone else's eyes. How else could a man, for example, imagine the constraints of pregnancy? In her polemical essay, 'A Defense of Abortion,' Judith Jarvis Thomson early on concedes to two apparently antithetical premises in the debate over abortion: (1) 'the fetus is a person from the moment of conception,' and (2) 'the mother has a right to decide what shall happen in and to her body' (p. 122). Having granted to each side of the issue its most central tenet, Thomson asks us to consider a scenario:

> But now let me ask you to imagine this. You wake up in the morning and find yourself back to back in bed with an unconscious violinist. A famous unconscious violinist. He has been found to have a fatal kidney ailment, and the Society of Music Lovers has canvassed all the available medical records and found that you alone have the right blood type to help. They have therefore kidnapped you, and last night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract poisons from his blood as well as your own. The director of the hospital now tells you, 'Look, we're sorry the Society of Music Lovers did this to you – we would never have permitted it if we had known. But still, they did it, and the violinist now is plugged into you. To unplug you would be to kill him. But never mind, it's only for nine months. By then he will have recovered from his ailment, and can safely be unplugged from you.' Is it morally incumbent on you to accede to this situation (p. 123)?

Through this example Thomson concretizes the two opposing premises and uses the obvious parallels to pregnancy to develop a distinction between killing a fetus and killing it unjustly. Her conclusion is that abortion is killing but not unjust killing.

My interest in Thomson is not the issue of abortion but how she engages the reader to grapple with it. She offers a hypothetical example to give substance to a general principle. But her example of the unconscious violinist is not merely an imaginary situation; it has certain characteristics and expectations that enable it to engage the readers and encourage them to actively participate in the issue and not just passively observe. Fictional examples that engage the audience in this way are known as thought experiments. In this paper I will (1) stipulate some criteria for what I will count as a thought experiment, (2) justify their choice as an object of study, (3) identify a significant mechanism in their presentation, (4) describe a method of analysis, (5) offer samples of analysis, and (6) offer some conclusions from this analysis for the study of thought experiments in particular and argument in general.

THOUGHT EXPERIMENTS DEFINED

Thought experiments are made-up scenarios, but their fictional nature should not relegate them to the realm of literature. Many critics and com-

mentators of thought experiments characterize them as arguments. Irvine in describing the value of thought experiments in the physical sciences says, 'thought experiments are to be understood as arguments concerning particular hypothetical events or states of affairs' (p. 158). Others like Norton go to great lengths to plot the argument in a thought experiment in a formally deductive form with premises and conclusion. He defends this practice: 'thus we analyze and appraise thought experiments by reconstructing them explicitly as arguments and testing them against just those standards which we apply to arguments of other forms' (p. 142). Hofstadter says candidly, 'On other occasions thought experiments, however systematically developed, are intended merely to illustrate and enliven ideas, and sometimes the boundaries between proof, persuasion, and pedagogy cannot be drawn' (p. 459).

Beyond the fact that they are a form of argument, however, there is not much precision in the literature about the characteristics of thought experiments. Most definitions of the term are couched in the experiment metaphor. Brown says, 'Thought experiments are performed in the laboratory of the mind. Beyond that metaphor it's hard to say just what they are' (p. 1). Sorensen says, 'A thought experiment is an experiment that purports to achieve its aim without the benefit of execution' (p. 205). Wilkes tries to depart from the metaphor a bit, saying that thought experiments are 'forays of the imagination . . . for concluding that a philosophical thesis is plausible or implausible . . . [which] obey many of the constraints on experimentation' (p. 2).

Some philosophical thought experiments, in fact, are elaborations of actual experiments. Derek Parfit, for example, in his attempt to refute the necessity of the unity of consciousness, offers an invented example built around Sperry's experiments with commissurotomy.

> . . . suppose that I have been equipped with some device that can block communication between my hemispheres. Since this device is connected to my eyebrows, it is under my control. By raising an eyebrow I can divide my mind. In each half of my divided mind I can then, by lowering an eyebrow, reunite my mind. This ability would have many uses. Consider [that] I am taking an exam, and have only fifteen minutes left in which to answer the last question. It occurs to me that there are two ways of tackling this question. I am unsure which is more likely to succeed. I therefore decide to divide my mind for ten minutes, to work in each half of my mind on one of the two calculations, and then to reunite my mind to write a fair copy of the best result (p. 246).

Parfit says of such examples '[their] impossibility is merely technical.' He believes that given enough time and resources such an experiment could be conducted. Nevertheless, such an example seems to cross a line beyond which there is no longer any basis in empirical observation. Philosophical thought experiments like Thomson's and Parfit's illustrate the kind of example I am interested in. They are counterfactual examples that bend the rules (humans' or nature's) in some important way.

THOUGHT EXPERIMENTS AS AN OBJECT OF ANALYSIS

Thought experiments occur frequently in the fields of personal identity, philosophy of mind, and ethics. Philosophical thought experiments seem worthy of scholarly inquiry for at least two reasons. First, they are commonplace. Brooks notes, 'Thought experiments seem too integral to philosophy since Plato's cave for us to abandon them' (p. 71). Goodin says, 'The methodological hallmark of *Philosophy & Public Affairs* is the "thought experiment"' (p. 8). Second, their use often evokes immediate and trenchant criticism. Barry complains about philosophers of social and political theory who rely on 'their one-sided diet of desert-island examples' (p. 3). Chandra criticizes thought experiments for being performed 'under the garb of scientific discoveries' (p. vii). Nevertheless, it seems that philosophers find it hard to resist offering and responding to thought experiments, perhaps because, as Jackson observes, the thought experiment 'provides ample scope for the exercise of ingenuity in the usually arid confines of academic writing' (Jackson, 1992, p. 528). In spite of their salience thought experiments have not received a lot of in depth and exclusive study, especially from the field of rhetoric.

THE MECHANISM OF THOUGHT EXPERIMENTS IN DIALOGUE

Most thought experiments are fairly well developed so as to flesh out for the reader an easily imaginable scenario. Some, in fact, are quite protracted, such as Gloria Steinem's essay, 'If Men Could Menstruate.' From the premise suggested by the antecedent clause of the title, Steinem develops four pages of consequents beginning with:

> Clearly, menstruation would become an enviable, boast-worthy, masculine event: Men would brag about how long and how much. Young boys would talk about it as the envied beginning of manhood. Gifts, religious ceremonies, family dinners, and stag parties would mark the day (p. 338).

Thought experiments, however, needn't be well developed or occupy the context of a formal argument complete with explicit premises and conclusion. Such one-line hypotheticals as Hare's 'We might think differently of abortion if wombs were made of glass,' or MacKinnon's 'If Blacks owned Black slaves, would that express white supremacist structure or subvert it?' (p. 119) both seem to conform to the more fully developed and extended examples like Thomson's.

Philosophical thought experiments often provoke responses in kind and evolve through a series of revisions as they pass back and forth between interlocutors. Perelman's notion of the emended analogy (p. 388) seems adaptable to describing what happens when, for example, Jackson substitutes Desmond Tutu for the famous violinist in Thomson's thought exper-

iment. Perelman notes that philosophers will often attempt to refute analogies by altering them in some way. He cites the example of Leibniz, who in refuting Locke's strict empiricism as expressed by the *tabula rasa*, or blank marble, analogy suggested that there might be 'veins in the block' corresponding to rough, unpolished innate ideas.

Similar moves can be seen in the responses to thought experiments. Searle's Chinese room, for example, starts out as a man in a room with a pile of Chinese symbols and a translation manual:

> Suppose that I'm locked in a room and given a large batch of Chinese writing. Suppose furthermore (as is indeed the case) that I know no Chinese, either written or spoken, and that I'm not even confident that I could recognize Chinese writing as Chinese writing distinct from, say, Japanese writing or meaningless squiggles. To me, Chinese writing is just so many meaningless squiggles. Now suppose further that after this first batch of Chinese writing I am given a second batch of Chinese script together with a set of rules for correlating the second batch with the first batch. The rules are in English, and I understand these rules as well as any other native speaker of English. They enable me to correlate one set of formal symbols with another set of formal symbols, and all that 'formal' means here is that I can identify the symbols entirely by their shapes. Now suppose also that I am given a third batch of Chinese symbols together with some instructions, again in English, that enable me to correlate elements of this third batch with the first two batches, and these rules instruct me how to give back certain Chinese symbols with certain sorts of shapes in response to certain sorts of shapes given me in the third batch. Unknown to me, the people who are giving me all of these symbols call the first batch a 'script,' they call the second batch a 'story,' and they call the third batch 'questions.' Furthermore, they call the symbols I give them back in response to the third batch 'answers to the questions,' and the set of rules in English that they gave me, they call the 'program.' Now just to complicate the story a little, imagine that these people also give me stories in English, which I understand, and they then ask me questions in English about these stories, and I give them back answers in English. Suppose also that after a while I get so good at following the instructions for manipulating the Chinese symbols and the programmers get so good at writing the programs that from the external point of view – that is, from the point of view of somebody outside the room in which I am locked – my answers to the questions are absolutely indistinguishable from those of native Chinese speakers (in Hofstadter and Dennett, 1981, p. 355).

In response to the system reply, which argues that even though Searle's demon in the Chinese room may not understand Chinese, the combined system of the room and all its contents including the demon do understand, Searle replies with what might be called the camera-phage variation of the Chinese room: 'Let the individual internalize all these elements of the system. . . . The individual then incorporates the entire system' (p. 359). In spite of devouring the whole room the individual, Searle would argue, is still in no better position to understand Chinese.

Another reply to Searle argues that the Chinese room system and by analogy the computer system it simulates would understand Chinese if the causal theory of reference were invoked whereby some causal connection between formal symbols in the program and referents in the world could be effected. To this reply Searle fashions what might be called the egg-foo-yung variation on his Chinese room thought experiment. He says, 'Suppose,

for example, that the symbol for egg foo yung in the Chinese room is actually causally connected to egg foo yung. Still, the man in the room has no way of knowing that. For him it remains an uninterpreted symbol, with no semantic content whatever' (1980, p. 452). Thus in response to criticism of the original thought experiment an emended version emerged. It is through this process of emendation that I wish to study thought experiments.

METHOD OF ANALYSIS OF THOUGHT EXPERIMENTS

Although thought experiments occur frequently in the fields of personal identity, philosophy of mind, and ethics, where a backdrop of analytic philosophy has made argumentative rigor a touchstone, many commentators have taken a non-propositional approach and observed a story-like quality in these accounts. Brown (1992) characterizes the thought experiment as 'a narrative explanation' (p. 275). Gooding (1992) regards thought experiments as 'process-narratives to which visualization is essential' (p. 281). Shapin (1984) notes an antecedent sense of narrative in the history of real experiments; he believes that Boyle's reports of his experiments were strongly narrative in order to allow readers to witness the experiment *in absentia*. Nersessian (1992) sees a parallel between this narrative character in real experiments and the presentation of thought experiments, a process she calls 'virtual witnessing' (p. 298).

Others have elaborated on the view of thought experiments as narratives but then gone on to discount its importance. Nersessian (1992), for example, explains how the thought experiment is a way for readers to simulate a real situation in their minds and notes, 'the narrative form of presentation plays a central role in communicating a thought experiment' (p. 292). But she then minimizes the role of language by claiming that the execution of this simulation is not on linguistic representations. For Nersessian the narrative details are part of the presentation of the thought experiment, but the reader's simulation of it is ultimately nonlinguistic.

Hacking (1992) also acknowledges but discounts the importance of the narrative aspect of thought experiments. In his response to Brown, Nersessian, and Gooding, 'Do Thought Experiments Have a Life of Their Own?' he equates a thought experiment to 'acting a part in a play' (1992, p. 307), where each reading voices the same dialogue and reaches the same conclusions. In fact, Hacking prefers to analogize thought experiments to jokes rather than to stories because they are 'well worked out items with a punch line' (p. 307). In short, Hacking believes thought experiments are static and immutable; they do not live but serve only to replay a prerecorded message.

But in denying life to thought experiments these latter commentators seem to have missed an opportunity to observe how the narrative aspects

of thought experiments have implications for the process whereby one version of a thought experiment can spawn another. In fact, this analysis suggests that each reading voices a different point of view to reach the same conclusions in some cases but different conclusions in others.

Narrativity has already been entertained as an alternative to the modernist way of knowing. Painting this alternative in broad strokes, Bruner (1991) notes that narrative ways of knowing have shown themselves to be antidotes to the shortcomings of the rationalist-empiricist program. Bruner has long maintained: 'we organize our experience and our memory of human happenings mainly in the form of narrative' (1991, p. 4). Given its universal necessity in the mental life of humans, he offers narrative as an 'instrument of mind in the construction of reality' (p. 6).

Sorensen (1992) suggests some narrative elements of thought experiments that might provide the kind of details that reflect justificatory practices when he notes: 'There is a difference between what the characters within a thought experiment can know and what the audience can know' (p. 273). This observation suggests the portrayal of the various characters both inside and outside the thought experiment may be relevant details in the presentation. The portrayal of the character of the narrator is one aspect of the Chinese room thought experiment that varies widely. This character represents the story's point of view and is expressed in the variations of the Chinese room thought experiment by the person of the verbs used in the narration of the action. Whether first, second, or third person, singular or plural, the narration of the Chinese room seems to reflect the author's philosophical position.

Also since thought experiments in general call on the reader to walk in another's shoes, they ask the reader to adapt to the action of the scene and his or her personal relationship to this action. This regard for the audience is evident in the relationships between the author of the thought experiment and the reader. Concerning these characters one might ask: 'How are the storyteller and the audience of the story represented?'

Of course the characters inside the Chinese room are no less important than the narrator and the reader on the outside. Their portrayal, in fact, may provoke their own biases to be wary of. Unger (1990), for example, feels that we are so used to the idea of character continuity in a story that when we see, for example, Captain Kirk of *Star Trek* enter one end of a teleporter we assume without question that the character emerging from the other end is the same Captain Kirk. Story-telling conventions seem to override doubts about personal identity. The characters inside the Chinese Room are also portrayed in a myriad of ways. One sees in the variations a veritable revolving door where a range of agents take turns at playing the role of symbol manipulator in the Chinese room. Concerning the characters inside of the Chinese room thought experiment one might ask: 'How are the characters in the story represented?'

ANALYSIS #1: SEARLE'S CHINESE ROOM

John Searle in a response to the first sally of criticism of his original account of the Chinese room cautions the reader of his version and any variation: 'Remember, in these discussions, always insist on the first person point of view' (Searle, 1980, p. 451). Most of the commentaries and variations have not taken this advice. A handful do maintain first person verbs in recounting and emending Searle's original. A few are in the second person. But most accounts are rendered in the third person. This mode of narrative would seem merely consistent with scientific and scholarly style guides were it not also compatible with certain philosophical ideologies. Regardless of which person tells the story, however, his or her identity (it's seldom feminine, however) through the pronouns reflects certain assumptions the author is holding about how the mapping of reader to agent in the room should occur and what its significance is.

In thought experiments, it seems, the reader is never far away from the author's awareness. To support his belief that consciousness is a matter of continuity of memory John Locke speculates:

> Had I the same consciousness that I saw the ark and Noah's flood, as that I saw an overflowing of the Thames last winter, or as that I write now; I could no more doubt that I who write this now, that saw the Thames overflowed last winter, and that viewed the flood at the general deluge, was the same self . . . (Locke, 1975, p. 45).

Locke's first-person report here is a practice typical of British empiricists and perhaps imitative of Descartes. During the author's account of his own introspection, the reader is expected to replicate the author's line of thinking and to arrive at the same conclusions. In short, the author's 'I' becomes the reader's 'I'. This expectation is not merely an undisguised rhetorical device but a way to endorse introspection as the proper way of studying the phenomenon of consciousness.

Searle's account of the Chinese room thought experiment seems to revert to Locke's technique of presentation, for it starts out as the first-person report of a man in a room with a pile of Chinese symbols and a translation manual:

> Suppose that I'm locked in a room and given a large batch of Chinese writing. Suppose furthermore (as is indeed the case) that I know no Chinese, either written or spoken, and that I'm not even confident that I could recognize Chinese writing as Chinese writing distinct from, say, Japanese writing or meaningless squiggles (Searle, 1980, p. 355).

While it is true that Searle's prose style has a reputation for being casual, maintaining as he does here an informal, first-person stance, his choice of first person verbs in the Chinese room example is, I believe, more than just a matter of consistency of style. The emphasis on the first person singular seems central to Searle's argument. Elsewhere in a related discussion Searle (1994) makes a case for the subjective nature of consciousness. The upshot of this subjectivity is that the mental is 'an

irreducibly first person ontology' (p. 95). As a result consciousness is not accessible to all observers in the same way. So by keeping the account of the Chinese room in the first person, Searle keeps the subjective aspect of the mental implicit and thereby encourages the reader to do likewise by running the simulation in his own mind. Searle's purpose here seems to be to segregate consciousness from other phenomena with respect to the way they can be studied, where the Cartesian distinction between *res cogitans* and *res extensa* is usually assumed. A first-person point of view serves to reinforce the subjective nature of the mental and so confers a certain epistemic privilege on itself.

Searle is quite consistent about this point of view when he offers or responds to any accounts of the Chinese room thought experiment. In questioning the basis of the systems response, he maintains, 'The only motivation for saying there *must* be a subsystem in me that understands Chinese is that I have a program and I can pass the Turing test' (Hofstadter and Dennett, 1981, p. 360). In his response to the robot reply he maintains 'I am the robot's homunculus, but . . . I don't know what's going on' (p. 363).

Dennett has modified the Chinese room thought experiment many times but never uses first-person pronouns to refer to the agent in the room. He works hard to embellish the story's details but always maintains a third-person point of view. Here is his version of what Searle calls the systems reply:

> Searle, laboring in the Chinese Room, does not understand Chinese, but he is not alone in the room. There is also the System, CR, and it is to that self that we should attribute any understanding . . . (Dennett, 1991, p. 439).

Here is his version of the robot reply:

> We have got Searle, a Chinese-speaking robot, who has incorporated the whole system in himself. Suppose I encounter him in the street and say: 'Hello, John, how are you?' What does your imagination tell you that happens? (Dennett, 1995, p. 66)

The shift from Searle's first-person account to Dennett's third-person seems to reflect their prior and broader differences over the issue of intentionality.

Intentionality is a philosophical term coming from the Latin *intendere*, to point, so agents with intentionality can have mental states that point to or are about something else outside themselves. Intentional states are about something: a man can think about a rock but not vice versa; a book can be about a rock but a rock cannot be about something. In fact, Bretano and Husserl have held up the intentional as the distinguishing feature between the mental and the physical.

Apart from these general notions about intentionality there is much disagreement and confusion about the specifics of this concept and indeed about its very validity. For example, intentional states in the context of philosophy generally and in the Chinese room thought experiment particularly are not to be identified with the everyday notions of intention and inten-

tional as something willed or desired. The latter are, in fact, special cases of intentional states.

Another distinction concerns types of intentionality. Searle (1983), for example, posits two types: intrinsic and derivative. Intrinsic intentionality is the real flesh-and-blood aboutness that occurs in living organic brains. Derivative intentionality is aboutness that depends upon an agent with the intrinsic type. A computer programmed to play chess has intentionality that is derived from the programmer. For Searle intrinsic intentionality is primary because it sets its own conditions of satisfaction, which includes a certain 'causal self-referentiality' (Searle, 1983, p. 49). This feature is meant to point out the weakness of causal theories of reference that claim meaning isn't in the head. Searle contends that in addition to a certain state of affairs in the world, meaning also requires an experience of meaning 'which is the embodiment or realization of the Intentional content' (p. 48). The conditions of satisfaction for an intentional state are not met until there is also an embodied experience of that state. It is this self-referentiality that necessitates for Searle a first-person account of intentional contents in general and a first-person account of the Chinese room thought experiment in particular.

Dennett, on the other hand, views intentionality in terms of the third person. He calls his view an intentional system theory as if to abstract intentionality from human tissue. His theory focuses on 'the performance specifications of believers while remaining silent on how the systems are to be implemented' (Dennett, 1987, p. 59). His is a behaviorist methodology, which presumes only 'the objective, materialistic, third-person world of the physical sciences' (p. 5). He contrasts his view with Searle's, which he calls 'a misconceived attempt to capture the internal point of view of a conscious agent' (Dennett, 1980, p. 430). Hence, Dennett is unwilling to let himself enter the Chinese room and experience it from the inside.

What do the presentations of these variations of the Chinese room thought experiment reveal about their presenters? First, the shifts in point of view seem to reflect commitments about the nature of the proper methodological approach to studying mind. Introspection, in particular, takes a beating. This traditional approach to mind represents one half of the asymmetry that British empiricists like Locke assumed between first-person and third-person approaches to knowledge. They believed that each of us had special access to our own minds in a way that was denied to others. For them introspection was undeniable and infallible. Taken to its extreme via Cartesian radical doubt, this doctrine makes the mind a better knower of itself than of the external world.

Most of Searle's opponents would distance themselves from such a view lest they be drawn into a spiral of solipsism. Introspection for them necessarily isolates one from others. This condition forces one to choose between the two mutually exclusive positions: (1) I'm the only conscious being in the universe and (2) All beings in the universe are conscious. In

either case no one's report suffices for proof of consciousness. Thus most functionalists avoid first-person accounts of the Chinese room thought experiment.

Searle would resent their characterizations of his position. He, in fact, would say to the contrary that he denies the infallibility of introspection, admitting that self-deception is possible. Nevertheless, he feels that an awareness of one's self is an essential feature of consciousness. Moreover, this self-consciousness is essentially subjective, a feature by which he salvages from the doctrine of introspection a privilege for first-person accounts.

However, functionalists like Dennett do not trust first-person accounts because they produce a 'cozy complicity' (Dennett, 1991, p. 70) between the writer and the reader such that the account becomes hopelessly fallible and incorrigible. Like Ryle, his mentor, Dennett believes we know ourselves and others via observable behavior, which, Dennett believes, should be recorded in the third person.

Functionalists, like Dennett, see Searle's insistence on first-person accounts as a commitment to introspection, which smacks of Cartesian dualism. If something is so subjective as to be untouched by third-person objectivity, it must be dualistic in nature. Functionalists believe that mysteries of the mind will surrender to a third-person objectivity just as the mysteries of other complex physical phenomena have. Such is the case with intentionality, which for Dennett comes in only one flavor – derived. In fact, by insisting that human brains are capable of only derived intentionality and that nothing has intrinsic intentionality Dennett seems to remove the issue of meaning from the conversation about human consciousness altogether.

In his first commentary on the Chinese room thought experiment Dennett (1980) contrasts his position with Searle's by noting, 'For Searle intentionality is rather like a wonderful substance secreted by the brain the way the pancreas secretes insulin' (p. 429). Searle (1994) would agree with this metaphorical characterization; he says explicitly of his theory of mind, 'Mental events and processes are as much part of our biological natural history as digestion, mitosis, meiosis, or enzyme secretion' (p. 1). For Searle the subjective sense of self is closely tied to an awareness of a particular body; in other words, consciousness is necessarily embodied. For Dennett the sense of self is abstractable from and independent of the body.

ANALYSIS #2: THOMSON'S UNCONSCIOUS VIOLINIST

In addition to the Chinese room thought experiment the literature of philosophy contains many other thought experiments that have spawned variations. By examining them in dialogic pairs, we might see how one

version of a narrative evolves from another, enhancing and mutating details for a particular argumentative effect. In each case the details selected reflect certain goals and self images, and all the while the reader is implicitly asked to try on this or that narrative.

Thomson's (1971) unconscious violinist is one other example that has provoked many such emended versions. In responding to Thomson's defense of abortion, Jackson (1992) in a subsequent article objected to the appeal to universals in Thomson's argument. He stated further, 'The point is made by working with Thomson's example' (p. 531) and offered this minor variation on her case:

> If instead of a sleeping violinist the beneficiary were Desmond Tutu, then I would not be surprised if some people in the victim's position might not think it right to accept the nine months wait. . . . Indeed, some of them might well organize the kind of kidnapping picture by Thomson, kidnapping both the sleeping woman and Tutu to save him against his own will. I would be much less sure than Judith Jarvis Thomson that it would be wrong to save Bishop Tutu at the cost of enslaving a woman's body for nine months (Jackson, 1992, pp. 531–532).

Thomson's violinist and Jackson's Tutu variation seem to illustrate intuitive forces pulling the reader in opposite directions. Thomson wants the reader to believe that pulling the plug on the violinist is justifiable homicide; Jackson replaces the violinist with the character of Desmond Tutu and calls it murder. Thomson (1971) acknowledges the role of intuition in adopting her belief, saying:

> If anything in the world is true, it is that you do not commit murder, you do not do what is impermissible, if you reach around your back and unplug yourself from that violinist to save your life (p. 52).

Jackson (1992) likewise acknowledges an appeal to intuition in deciding this case, saying, '[the beneficiary's] character counts no less than rules and principles' (p. 531). But what about the beneficiary's identity changes the force of the analogy? I believe it is the relationship created between the beneficiary and the victim that serves to endear the reader in one case and to alienate in the other. This relationship reveals itself through the narrative details used by Thomson and Jackson.

Thomson's account of the beneficiary seems marked by greater abstraction than Jackson's. Thomson (1971) asks the reader to imagine '*an* unconscious violinist.' Jackson (1992) invokes a specific person, Desmond Tutu. Thomson's violinist is associated with a fictitious 'Society of Music Lovers;' Jackson relates Desmond Tutu to the real context of 'South African politics and religion.' When it comes to describing the victim, however, Thomson and Jackson seem to reverse their uses of the abstract. Thomson posits the most concrete identity possible for the victim – the very person of the reader, saying, 'You wake up in the morning and find yourself back to back in bed with an unconscious violinist' (p. 51). Jackson (1992) refers to the victim vaguely as 'a sleeping woman' (p. 531).

Their contrasting modes of description seem to reflect their regard for the beneficiary and the victim. Thomson's argument presupposes greater concern for the victim than for the beneficiary; for this she portrays the former in concrete terms, the latter in abstract. Jackson presumes greater concern for the beneficiary, which he portrays concretely, than for the victim, which he portrays abstractly. Such moves are common rhetorical ploys: as Perelman (1969) in his discussion of presence notes, 'To create emotion, it is essential to be specific' (p. 147).

Many other mutations of the original unconscious violinist thought experiment appear in the philosophical arguments over abortion. Often the emended thought experiment will shift the position of the narrator, or what becomes the position of the reader, as he or she simulates the imagined scenario. In the original the victim (and so the reader) has been kidnapped; in Warren's (1973) variation the victim volunteers to help the violinist via a lottery, while in Davis's (1983) variation the victim volunteers but changes her mind. In the original we, the readers, are asked to imagine being the kidnapped victim; in Wreen's (1992) variation we are asked to imagine being the violinist.

Some variations alter certain details of the scenario. Thomson's violinist is unconscious; Beckwith's (1992) is conscious. Thomson has the victim unplug the violinist; Finnis (1973) has a bystander unplug the violinist. Thomson's violinist is unrelated to the victim; Wennberg's (1985) violinist is the victim's son.

Other variations alter some philosophical underpinning in the case. Thomson forces a dilemma on the reader: either you stay with or leave the violinist; Wreen (1992) suggests that there are other courses of action. Thomson's account is based on the premise that moral obligations are voluntary; Beckwith (1992) feels that some moral obligations accrue naturally. The same sort of opportunities for analysis arise in the case of the violinist as were carried out on the man in the Chinese room. Both cases provide fertile ground for exploring dialogical impasse. An analysis of the variations of the unconscious violinist thought experiment might clarify the nature of the impasse in the controversy over abortion and perhaps identify overlapping perspectives.

CONCLUSIONS

In this paper I have argued that thought experiments are characterized by a reciprocal influence between the claims they support and the details they embody, as if the thought experimenters themselves were caught in a hermeneutic circle. My route to this conclusion has included these points: (1) thought experiments are distinctive argument moves, (2) their details become significant when they are subject to the process of emendation, (3) emended thought experiments are amenable to a non-propositional,

rhetorical analysis, and (4) such analysis suggests a reciprocal influence between the claims and details of thought experiments.

Thought experiments are contrived, counterfactual examples that do not or could not exist. In spite of, or perhaps because of this feature, their power to lure readers into another world is compelling. The setting of the Chinese room or the plight of the unconscious violinist seems irresistable. Though readers are easily drawn to entertain the experience of such an imagined scenario, their assent to its intended conclusion is not foregone. Readers of a thought experiment in support of abortion rights can reconstruct it to oppose abortion by changing merely one detail – the identity of the main character. The dialectic that emerges from such an exchange of emended thought experiments, I have tried to show, is characterized by a reciprocal influence between the assumptions of the arguer's position and the details that describe the thought experiment. Dennett's insistence on third person point of view in the Chinese room both supports his functionalist conclusion and reflects his functionalist assumptions. This analysis of narrative details has acknowledged the propositional content of the controversies surrounding a given thought experiment, but it has also attempted to preserve the significance of the details – a feature that might otherwise be lost in an analysis that reduces arguments to premises and conclusions.

REFERENCES

American Cancer Society: n.d., poster, 91-100M-No5648.

Aristotle: 1984, *Rhetoric* (W. Rhys Roberts, Trans.), The Modern Library, New York.

Barry, Brian: 1979, 'On Editing Ethics', *Ethics* **90**(1), 1–6.

Beckwith, F. J.: 1992, 'Personal Bodily Rights, Abortion, and Unplugging the Violinist', *International Philosophical Quarterly* **32**(1), 105–118.

Brooks, D. H. M.: 1994, 'The Method of Thought Experiment', *Metaphilosophy* **25**(1), 71–83.

Brown, J. R.: 1992, 'Why Empiricism Won't Work', *PSA 1992* **2**, 271–279.

Brown, J. R.: 1991, *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*, Routledge, London.

Bruner, J.: 1991, 'The Narrative Construction of Reality', *Critical Inquiry* **18**(1), 1–21.

Chandra, S.: 1977, *Identity and Thought Experiments*, Institute of Advanced Study, Simla.

Davis, M.: 1983, 'Foetuses, Famous Violinists, and the Right to Continued Aid', *Philosophical Quarterly* **33**(132), 259–278.

Dennett, D. C.: 1995, 'In Defense of AI', in Peter Baumgartner (ed.), *Speaking Minds: Interviews with Twenty Eminent Cognitive Scientists*, Princeton University Press, Princeton, pp. 57–69.

Dennett, D. C.: 1991, *Consciousness Explained*, Little Brown, Boston.

Dennett, D. C.: 1987, *The Intentional Stance*, The MIT Press, Cambridge, MA.

Dennett, D. C.: 1980, 'The Milk of Human Intentionality', *The Behavioral and Brain Sciences* **3**, 428–430.

Finnis, J.: 1973, 'The Rights and Wrongs of Abortion: A Reply to Thomson', *Philosophy and Public Affairs* **2**(2), 135–145.

Goodin, R. E.: 1982, *Political Theory and Public Policy*, The University of Chicago Press, Chicago.

Gooding, D. C.: 1992, 'What is Experimental about Thought Experiments?', *PSA 1992* **2**, 280–290.
Hacking, I.: 1992, 'Do Thought Experiments Have a Life of Their Own? *PSA 1992* **2**, 302–308.
Hare, R. M.: 1981, *Moral Thinking*, Clarendon Press, Oxford.
Hofstadter, D. R. and D. C. Dennett (eds.): 1981, *The Mind's I: Fantasies and Reflections of Self and Soul*, Bantam Books, Toronto.
Irvine, A. D.: 1991, 'Thought Experiments in Scientific Reasoning', in T. Horowitz and G. J. Massey (eds.), *Thought Experiments in Science and Philosophy*, Rowman & Littlefield Publishers, Inc., Savage, MD, pp. 149–165.
Jackson, M. W.: 1992, 'The *Gedankenexperiment* method of Ethics', *The Journal of Value Inquiry* **26**, 525–535.
Locke, J.: 1975, 'Of Identity and Diversity', in J. Perry (ed.), *Personal Identity*, University of California Press, Berkeley, pp. 33–52.
MacKinnon, C. A.: 1989, *Toward a Feminist Theory of State*, Harvard University Press, Cambridge.
Nersessian, N. J.: 1992, 'In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling', *PSA* **2**, 291–301.
Norton, J.: 1991, 'Thought Experiments in Einstein's Work', in T. Horowitz and G. J. Massey (eds.), *Thought Experiments in Science and Philosophy*, Rowman & Littlefield Publishers, Inc., Savage, MD, pp. 129–148.
Parfit, D.: 1984, *Reasons and Persons*, Clarendon Press, Oxford.
Perelman, C. and L. Olbrechts-Tyteca: 1969, *The New Rhetoric*, University of Notre Dame Press, Notre Dame.
Searle, J. R.: 1994, *The Rediscovery of Mind*, MIT Press, Cambridge.
Searle, J. R.: 1983, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge.
Searle, J. R.: 1980, 'Minds, Brains, and Programs', *The Behavioral and Brain Sciences* **3**, 451–455.
Shapin, S.: 1984, 'Pump and Circumstance: Robert Boyle's Literary Technology', *Social Studies of Science* **14**, 481–520.
Sorensen, R. A.: 1992, *Thought Experiments*, Oxford University Press, Oxford.
Steinem, G.: 1983, *Outrageous Acts and Everyday Rebellions*, Holt, Reinhart, and Winston, New York.
Thomson, J. J.: 1971, 'A Defense of Abortion', *Philosophy and Public Affairs* **1**(1), 47–66.
Unger, P.: 1990, *Identity, Consciousness, and Value*, Oxford University Press, New York.
Warren, M. A.: 1973, 'On the Moral and Legal Status of Abortion', *Monist* **57**(1), 43–61.
Wennberg, R.: 1985, *Life in the Balance: Exploring the Abortion Controversy*, Eerdmans, Grand Rapids, MI.
Wilkes, K. V.: 1988, *Real People: Personal Identity Without Thought Experiments*, Clarendon Press, Oxford.
Wreen, M.: 1992, 'Abortion and Pregnancy Due to Rape', *Philosophia, Philosophical Quarterly of Israel* **21**(3 and 4), 201–220.