Do Biases in Thought Experiments Render Experimental Philosophy Useless?

Amanda Cohen

Philosophy 550: Senior Project

Advisor: Professor Janack

Hamilton College

Fall 2012

Do Biases in Thought Experiments Render Experimental Philosophy Useless?

I. Introduction

With the emergence of experimental philosophy and the use of thought experiments, the line distinguishing philosophy from psychology has been blurred. With philosophical methodology encroaching upon that of the sciences, such as psychology and sociology, thought experiments should receive the same scrutiny that scientific experiments face. Empirical data can be useful for many findings, but if the experiment obtaining the data is not reliable, then the evidence should not be relied upon for determining philosophical conclusions. Much of the discussion surrounding the reliability of ex- phi, as experimental philosophy is called, is concerned with the trustworthiness of intuitions and whether or not intuitions can serve as reliable data and evidence for philosophy. However, we cannot address the question of whether or not intuitions are reliable until we examine the method used to obtain the intuitions.

In this paper I will scrutinize the methodology of how philosophers conduct thought experiments and expose biases that are inherent in this methodology. I will use Harry Frankfurt's thought experiment outlined in "Alternate Possibilities and Moral Responsibility" to show the inherent biases that this, and all other, experiments are susceptible to. I will also explain why these biases are both unavoidable and provide a serious threat to the reliability of using thought experiments to acquire evidence for philosophical conclusions. Finally, I will critique literature regarding the use of thought experiments and show that, despite

my arguments against thought experiments, this procedure should still be pursued by experimental philosophers because thought experiments allow philosophers to make advancements by looking at old topics from a new perspective.

II. The Framework of a Thought Experiment

A. Frankfurt's Thought Experiment:

Harry Frankfurt tries to get an idea of what humans think about moral responsibility in his paper "Alternate Possibilities and Moral Responsibility," which was published in December 1969. It is generally believed that people will not ascribe moral responsibility to an agent when he or she could not have done otherwise. Human intuitions generally show that people find agents morally responsible for their actions only when alternate possibilities were available to them. Frankfurt concludes that the principle of alternate possibilities should be amended because it is possible that an agent has decided to act in a certain way before he or she was coerced into acting in that same way. In other words, he wanted to commit the coerced action anyway or even before the coercion began. One is generally found to be morally responsible if an alternate possibility to his action was possible; however, he is not usually found morally responsible if there were alternate possibilities but he was coerced to act in the way that he did. Frankfurt shows that one might be coerced into an action after he had chosen to act in that same way, which would make him not morally responsible for an action he had previously decided to commit. Thus, Frankfurt proposes the following revision to the principle of alternate possibilities: "a person is not morally responsible for

what he has done if he did it only because he could not have done otherwise" (Frankfurt, 1969, 838). This differs from the original principle of alternate possibilities: "a person is morally responsible for what he has done only if he could not have done otherwise" (Frankfurt, 1969, 829). The difference lies in the "only because" addition to the principle. For example, according to the original principle of alternate possibilities Emma would be found morally responsible for stealing if she stole only because she didn't have another option. Frankfurt recognizes that Emma could have decided to steal before she realized that she didn't have any other option (in which case being coerced did not influence her decision); according to the amended principle, Emma would only be found morally responsible for stealing if it was a direct result of the coercion. Even though this finding was not necessarily based on what Frankfurt found by surveying intuitions, he made a significant impact on how people view moral responsibility by raising a new point about moral responsibility. Thus, he did not answer the question that he set out to answer using the methodology that he had intended (how people responded to his scenario), but he still made a worthwhile gain by using a thought experiment.

Frankfurt's thought experiment consisted of the following scenarios: in the first scenario, Frankfurt describes "Jones 1" as "not a reasonable man." This man does what he once decided to do no matter what happens after and no matter the cost. Thus, the threat exerted no effective force upon him; he acted without any regard to it and as if he were not aware of it. The situation did not involve any coercion, so the threat didn't lead Jones1 to do what he did nor did it prevent him

from doing otherwise. People generally conclude that the fact that he was threatened does not reduce moral responsibility for his action.

In the second scenario, Frankfurt describes "Jones 2," who was "stampeded by the threat." He would have performed the action regardless of what decision he had already made because the threat moved him so much that he completely forgot the decision he had already made and did what was demanded of him due to his fear of the penalty. The fact that he had already decided to perform that action is irrelevant because fear led him to act. People find that Jones 2 is not morally responsibly because he acted due to coercion.

Finally, Frankfurt describes "Jones 3," who was "neither stampeded by the threat nor indifferent to it." The threat resonated with him as it would with any reasonable man, but Jones 3 had decided to act in that way already. When he acted, he acted in accordance with the decision / the reasons he made before the threat. Jones3 would have made the same decisions that the threat led him to make even if he had he not already made the same decision anyway. If he had acted on the basis of his own decision, then his moral responsibility would be unaffected by the threat. However, since he would have submitted to the threat, he could not have avoided doing what he did. This example makes it hard to prove that coercion exists because Jones 3 acted as he would have if the threat had not existed, but, at the same time, he acted as a response to the threat. Thus, if we apply the principle of alternate possibilities, Jones 3 would not have been held morally responsible. But, as Frankfurt proposes, he would have still identified with the same action regardless of

the threat because he had made his decision prior to the threat. Thus it would seem that we should find that Jones 3 is morally responsible.

Frankfurt's thought experiment, like many thought experiments, consists of a hypothetical situation and three different ways the scenario plays out. The participant has to decide whether an imaginary person is morally responsible for his or her action. Frankfurt explains how we could imagine details that would make it reasonable for us to think that the person was (1) coerced to perform the action in question, (2) that he could not have done otherwise, and (3) that he bears no moral responsibility for having done what he did. The question then follows, what about the situation makes it so that the person is not morally responsible for his action?

Over the years, Frankfurt's thought experiment has received a lot of attention, both praise and criticism, and has also been used many times and rewritten by different philosophers. This is where the unavoidable danger of implementing Frankfurt's thought experiment comes in because slight variations in the wording of the thought experiment can evoke significantly different responses and results. Here is an example of an updated Frankfurt style case written by John Martin Fischer in "The Frankfurt Cases: The Moral of the Stores:"

> Because he dares to hope that the Democrats finally have a good chance of winning the White House, the benevolent but elderly neurosurgeon, Black, has come out of retirement to participate in yet another philosophical example. (After all, what would these thought-experiments be without the venerable ´eminence grise—or should it be noire?) He has secretly inserted a chip in Jones's brain that enables Black to monitor and control Jones's activities. Black can exercise this control through a sophisticated computer that he has programmed so that, among other things, it monitors Jones's voting behavior. If Jones were to show any inclination to vote for McCain (or, let us say, anyone other than Obama), then the computer, through the chip in Jones's brain, would intervene to assure that he actually decided to vote for Obama and does so vote. But if Jones decides on his own to vote for Obama

> (as Black, the old progressive would prefer), the computer does nothing but continue to monitor—without affecting—the goings-on in Jones's head. Now suppose that Jones decides to vote for Obama on his own, just as he would have if Black had not inserted the chip in his head. It seems, upon first thinking about this case, that Jones can be heldmorally responsible for his choice and act of voting for Obama, although he could not have chosen otherwise and he could not have done otherwise (316).

Both the names chosen and the situation described play a very large role in a person's response to the thought experiment. A handful of philosophers have addressed the idea that the way the experiment is presented can evoke various responses, rendering the thought experiment unreliable.

III. How the Wording of a Thought Experiment Biases the Results

A. Abstract Vs. Concrete Conditions:

One bias inherent in the construction of a thought experiment has to do with the way the scenario or question is presented. Nichols and Knobe (2007) found that when subjects are presented with an abstract question about whether agents in a deterministic universe are fully morally responsible, 86% of people believe that they are not completely morally responsible because they simply do what they are determined to do. On the other hand, when a shameful act is described in detail to a person (in the same deterministic universe), 72% of subjects believe that he *is* morally responsible. Essentially, the same question is asked twice, but one time with more details. Thus, the level of detail leads to two very different (and incompatible) responses (Nichols & Knobe, 2007, 670). Thus, we can conclude that the results depend so heavily on how the researcher phrases the thought experiment and the level of detail presented to the subject.

It is possible that a Frankfurt experiment could elicit completely different responses about moral responsibility depending on how abstract or how concrete the scenario is. So, for example, if the scenario describes Sarah robbing a bank with a gun a majority of people will conclude that Sarah is morally responsible. But, if we change the scenario and instead describe "a person" doing something illegal, people will be less inclined to ascribe moral responsibility to the actor. Even though both questions are trying to survey intuitions about the same principle (moral responsibility) the difference in the level of detail provided in the different scenarios will yield quite different data. Even though this finding by Nichols and Knobe is specific to moral responsibility (the same principle that Frankfurt's experiment tests), the same bias can plague other thought experiments and concepts just as easily. As Chris Weigel points out, "it seems highly implausible that people switch senses of moral responsibility when they switch from abstract to concrete descriptions of the same action" (Weigel, 2009, 235). The fact that people respond differently to different wording in the examples means that we cannot conclude that the responses represent the respondents' actual beliefs about moral responsibility, nor do they help us understand true beliefs about moral responsibility.

B. First Person Exceptionalism:

Our beliefs about moral responsibility also change whether or not we are involved in the situation. We hold ourselves to different standards than we hold others to. Tamar Gendler, in her paper *Philosophical Though Experiments, Intuitions, and Cognitive Equilibrium,* points out that we "consistently overestimate the

likelihood that [we] will act generously or selflessly, while accurately predicting the ungenerosity and selfishness of others (whom they most likely turn out to resemble)" (Gendler, 2007, 81). Thus, our thoughts about moral responsibility change depending on who is involved; we think more highly of ourselves than we do of others. If a Frankfurt case were to include the subject, or maybe even a loved one, then we will respond differently than if we were totally unbiased by the names or people in the scenario.

It may seem obvious that we would ascribe responsibility differently to ourselves compared to others, but we still have to take that difference into consideration when studying human attributions of moral responsibility. We cannot get an accurate picture about the way people think about moral responsibility if we do not factor in all circumstances in which these beliefs change. First person exceptionalism might be one of the easier biases to account for, but it still represents a threat to understanding the entire picture of how people ascribe moral responsibility. Despite how simple it seems to account for this bias, most philosophers still don't factor in first person exceptionalism when conducting and drawing conclusions from their thought experiments that test intuitions.

IV. How Differences in the Brain Affect Thought Experiments

A. The Affect of Stimulating Different Regions of the Brain:

One explanation for the fact that people respond differently to vividly described scenarios than they do to "thinner" scenarios might be that these different scenarios activate different parts of the brain. We will now consider that claim.

Depending on which region of the brain is activated, we produce different responses. If a Frankfurt experiment were framed to activate the "higher cognitive" region of the brain, then peoples' intuitive responses will be different compared to a thought experiment that tests the same condition, but activates the "emotional/ social regions" of the brain.

The famous trolley case reveals that people believe that it is morally permissible to intentionally turn the trolley so that it only kills one worker instead of allowing it to continue down its path, which would kill five workers. However, in a very similar case, which is known as the "fat man" case, people don't believe that it is morally permissible to push a fat man off a bridge to stop the trolley from killing five workers. In both instances the utilitarian calculation is the same: either one man dies or five men die but in the "fat man" case people don't believe we should sacrifice the one man while in the former case they think we should. Gendler explains these findings, "the original trolley case produces increased neural activity in "higher cognitive" regions of the brain, cases such as fat man (where the imagined action is "up close and personal") produce increased neural activity in "emotional/ social regions" (Gendler, 2007, 77). From this finding we can conclude that depending on which area of the brain happens to be stimulated by a scenario, the researcher will get different answers even to the same moral question.

Even though both the trolley thought experiment and the fat man thought experiment test a person's intuition regarding the same principle, whether it's morally acceptable to kill the one to save the many, the intuitive responses are overwhelmingly different. Thus, depending on which region of the brain is

stimulated by the thought experiment, a person's response changes. If this is true, then how can we get an accurate picture of human beliefs in general about the morality of a certain situation when answers vary so significantly based on how the phrasing of the thought experiment stimulates the brain? We cannot; therefore, we cannot draw an accurate conclusion about human intuitions about moral responsibility from thought experiments.

B. Some People are More Susceptible to Biasing Factors than Others:

There is a personality trait called Need for Cognition (NFC), which predisposes people to think differently about thought experiments. As tested by Cacioppo and Petty in 1982 and addressed by Weigel, this trait measures the degree to which a person "engages in and enjoys thinking" (Weigel, 2009, 226). Also, Weinberg, Alexander, and Gonnerman correlated scores on this measure with responses to thought experiments and found that a person's level of NFC influences how susceptible a person is to biasing factors- such as priming, framing, and other factors concerned with the phrasing of the question. They found that "mid and low NFC subjects are susceptible to priming" (Weigel, 2009, 226). For example, a person who falls in the mid-low NFC range, is found to be more susceptible to biasing factors, so whatever that person did preceding his participation in the thought experiment would have more of an influence on his response. On the other hand, high NFC individuals are also subject to priming, but the priming affects them in the reverse direction compared to the people with a lower NFC: "high NFC subjects seem to overcompensate for the default position" (Weigel, 2009, 226). Because the

high NFC participants are more mindful of the factors that could bias their response, they overthink the thought experiment, which still does not give the researcher an accurate idea of the participant's true intuition. Thus, even a personality trait such as a person's need for cognition can affect a person's response to a thought experiment.

This finding also threatens the reliability of thought experiments because if the researcher only surveys people who have either a high NFC or a low NFC without getting an equal sample of both, the results can be skewed. Experimental philosophers who obtain their participants by conveniently surveying a class or a particular group are not getting an accurate representation of the general population because people in these groupings might share a similar NFC level. Therefore, they are not getting an accurate representation of human belief.

V. How the Need to Uphold a Reputation or Self-Image Biases a Thought Experiment

A. Motivation to Control Prejudice as a Concern for Thought Experiments:

Self-Monitoring can also affect the ways that people respond to thought experiments. Plant and Devine are just two of the people who have outlined theories of contemporary prejudice, which claim that everyone is to some degree prejudiced even though not everyone is consciously aware of his or her own prejudices. As Grandall and Eshleman (2003) found, "because of the prejudice that affects them, people will sometimes feel an impulse to behave in a prejudiced or discriminatory manner but will restrain that behavior because the egalitarian aspect

of their value systems motivates them to act in an unprejudiced manner" (Whitley & Kite, 2010, 178). If we grant that there is a societal norm present in the United States that discourages the expression of prejudice, then it will follow that there is some reason to believe that people are motivated to comply with that norm. However, according to Plant and Devine, a person can be motivated to conform to a societal norm by either internal or external factors. Forsyth (2006) defines 'Societal norms' as "informal rules that groups develop that describe how to be a group member" (Whitley & Kite, 2010, 181). These rules govern both behaviors and attitudes. Internal motivations stem from a personal belief system in which, in this case, a person internalizes the belief that prejudice is wrong. Thus, people are internally motivated to act nonprejudiced do so because it is personally important to them. On the other hand, the external source that motivates people to comply with the nonprejudiced norm is a result of social pressure. These externally motivated people act in a nonprejudiced way in order to avoid negative reactions from other people.

As Devine explains, "people high in internal motivation to control prejudice… have so thoroughly integrated their nonprejudeced standards into their personal belief systems that they automatically control even implicit and physiological indicators of prejudice. Furthermore, people who are both high in internal motivation and low in external motivation show low levels of implicit prejudice when distracted, suggesting that they control their negative attitudes automatically, without conscious effort" (Whitley & Kite, 2010, 180). On the other hand, a person who is motivated by external motivation (like societal pressures) to control

prejudice is actually associated with exhibiting more prejudice. Plant and Devine (2001) hypothesize that this is because people who are high in external motivation and low in internal motivation probably try to avoid situations that make them interact with members of minority groups, which would make them have to avoid pressure from others to control their prejudices. They also report that they feel pressure to act "politically correct." Thus, societal pressure (or even the feeling of societal pressure) is biasing their true answers. Based on this finding, we can also conclude that people who are high in external motivation to control prejudice who participate in a thought experiment might feel uncomfortable about having to judge the actions of a person with a stereotypically black name, which could be because these subjects usually try to avoid interracial interactions.

The motivations for one subject to hide his or her prejudiced response is incredibly influenced by whether or not one's society and upbringing can play a large role in his or her response to thought experiments. Towles-Schwen & Fazio (2001) explain that people who attempt to avoid arguments tend to have been raised by prejudiced parents and have had few encounters with black people as a child; thus, media portrayals account for their main exposure to Black people (Whitley & Kite, 2010, 182). This means that a person's response to a thought experiment is influenced by his upbringing and his parents' level of prejudice, which biases the subject as data concerning intuitions, especially when names associated with a specific race are involved. Thus, if the philosopher is only looking to apply his or her results to one particular community or society, then the findings might accurately represent that community or society. However, if the philosopher is

looking to survey the general public regarding intuitions about moral responsibility, then these responses will not help in that pursuit. A person's response will change due to his or her motivation to control prejudice, which makes the results specific to a particular community and cannot be generalized to account for human intuitions overall. Another point demonstrated by this finding is that a person's upbringing will significantly alter how he views minorities and what kind of motivation, internal or external, he controls in his response. Thus, the researcher is not getting that person's "true" response to the question. As we can see, the type of motivation a person controls will also alter how he or she responds to a thought experiment. In terms of the Frankfurt example, a Frankfurt thought experiment cannot be a reliable source for obtaining intuitions about moral responsibility when the sample population only accounts for a specific community, especially one that is relatively homogenous in its view about race.

One might question why the factor of motivation to control a prejudice response renders the data unacceptable. If the thought experiment appears to have names typically associated with minorities, then people will not give their honest response either in fear of being judged or because they don't want to go against their internal belief system, if their intuitive response seems to be based on racial or ethnic prejudice. From this, one could draw the conclusion that if all of the names are associated with white people, then there will not be a need to control one's implicit reaction. However, this inference can be equally as biasing because the subject should not account for the differences in race while assessing one's views regarding moral responsibility.

This finding provides another threat for the use of thought experiments in philosophy because the fact that people can control their natural reactions so quickly makes us question whether or not we can ever get an accurate account of initial responses because there is so much processing that happens instantaneously. If a person has a gut reaction and is able to realize that the response is socially "wrong" and can also recognize another option that is more socially acceptable, he or she can alter his response all in a matter of a second. Thus, we cannot get an accurate picture of a person's intuition, especially if he can also control the physiological reactions associated with an implicit response.

Although this poses a worry for thought experiments in experimental philosophy, these findings do not threaten the value of the psychological studies because psychologists are content to simply document how people respond whereas philosophers seem to want to draw normative conclusions on the basis of these data. As Plant and Devine and many other psychologists have proven, psychologists can uncover the differences in people's responses. However, philosophers conducting thought experiments rarely attend to the ways in which responses to thought experiments are susceptible to biasing influences. Thus, this argument does not undermine the credibility of the psychological use of this data. However, the data does not provide an adequate foundation for creating larger conclusions about what people think about philosophical topics, such as moral responsibility.

Another related example further shows how revealing race in a thought experiment can influence a person's response. David Pizarro presented subjects

with a thought experiment similar to the fat man and the trolley cases, but he varied the nature of the sacrifice involved. In one scenario: "a man named Chip Ellsworth III could be thrown off a bridge to stop a trolley hurtling toward 100 members of the Harlem Jazz Orchestra; in the other, a man named Tyrone Peyton could be thrown off to save 100 members of the New York Philharmonic" (Gendler, 2007, 78). Chip and the Philharmonic represent white elite social and cultural values whereas Tyrone and the Harlem Jazz Orchestra are associated with the African American population. In these examples, people are more likely to sacrifice the white man than the black man; Pizarro concludes that this is an overcorrection to an initial instinctively racist response. Even though there is no moral difference between Chip and Tyrone, subjects tend to respond differently. But if we were to use these responses as data about how we ought to make moral decisions, then we would draw different conclusions for each scenario; the differences would be based on morally irrelevant aspects of the scenarios.

21st century Americans don't want others to think they are racist, so they will change or "overcorrect" their answers if the scenarios activate racist assumptions. Frankfurt's thought experiment is susceptible to this "overcorrection" because the names that are chosen for the thought experiment can sway a person's response by drawing on their intuitive assumptions about the race or ethnicity of the actors they are asked to judge. The phenomenon of "overcorrection" demonstrates how a person's response, even though it seems intuitive, is not necessarily a reliable picture of a person's true natural belief.

B. Bystander Effect:

A person will respond differently to a thought experiment depending on who is around him or her. If someone is nearby to observe a person's response, then there is a good chance that the person will answer differently compared to if the subject were alone. Even if the person is not actually being watched, but feels like he or she is being watched, then he will alter his response. Thus, a philosopher can obtain varying results from the same exact thought experiment depending on the way in which the experiment was implemented (i.e. whether the person was being watched while responding or whether he was completely alone). This bias is not concerned with the wording of a question or an encounter that a person experienced immediately preceding participation in the thought experiment, but rather is caused by an observer or the feeling of being watched.

As described by Buckwalter and Stich, Rigdon et al. (2009) shows how a person's behavior changes when he feels like he is being watched. Rigdon et al. conducted a study that demonstrates the effects of being watched (or even the sensation of being watched). In the study, 51 men and 62 women were presented with either the "face" or the "control" version of the task that the researchers were using. The face consisted of three dots that resembled two eyes and a mouth while the control had three dots that did not resemble a face. Participants were asked to write down how they would divide money ($10) between themselves and a recipient that they were paired with. Rigdon et al. (2009) found, "in the control condition 37% of men transferred $1.00 or more, but in the presence of a weak social cue (three dots arranged to look like a face) 79% of men transferred $1.00 or

more to the "Receiver" they had been paired with" (Buckwalter and Stich, 2010, 25). When a person felt like he was being watched he generally transferred more money to the "Receiver" compared to when he did not feel the presence of a "social cue."

This result exposes another bias facing thought experiments. If a Frankfurt case were proposed to a person with the philosopher, other participants, or any other individual in the room, then the participant might respond differently than if he or she were completely alone, with no sensation of being observed. From these results we can conclude again that we do not always follow our immediate responses.

V. Thought Experiment Responses are Subject to Personal Experience and are Temporal

A. Affects of Priming and Framing

Information that comes immediately before a question can also significantly influence a subject's response. Techniques of unconscious priming are extremely influential and, sometimes, uncontrollable. Gendler draws on another study done by Pizarro, which shows that when people were primed with patriotic terms they were more likely to assess American-on–Iraqi unintentional collateral damage to be morally acceptable than vice versa; however, when subjects were primed with multicultural terms, they were more critical of the same damage.

Another act of priming that can bias a thought experiment is when a philosopher's personal values are somehow accidentally (or not so accidentally) revealed before the subject has to respond to the thought experiment. If this

happens, the philosopher will get extremely biased results. Many times a philosopher's results will match that of his or her own personal belief, which is one of the main reasons that Gendler believes thought experiments can actually be used, not as information about moral beliefs, but rather as an instrument of persuasion. If the researcher believes that the individual in Frankfurt's thought experiment is morally responsible, then- it is possible- that his findings will reflect this personal belief. Also, a philosopher cannot always account for priming because something that happened to the subject before he or she answered the thought experiment can influence his or her response and cannot be accounted for by the researcher.

Another person who has addressed the issue of framing a thought experiment is Jonathan Weinberg. In his "Truetemp" thought experiment, he surveyed intuitions about whether subjects find a person, "Mr. Truetemp," possesses knowledge. From this thought experiment, he discovered (as Tamar Gendler points out), "(1) willingness to attribute knowledge in the Truetemp Case increases after being presented with a clear case of non-knowledge, and (2) willingness to attribute knowledge in the Truetemp Case decreases after being presented with a clear case of knowledge" (Gendler, 2007, 79). Thus, the framing of a question can sway a person's answer. This bias can be useful in understanding how framing questions influences a person's response, but in terms of conducting a thought experiment, framing poses a big threat to the success of the study because it is yet another thing that can influence a person's response.

One might think that framing is an easy problem to solve because all the philosopher has to do is be sure not to frame his or her thought experiment at all or,

if he wants or needs to frame the question, then he can account for the framing in the conclusion. For example, rather than concluding *x* about moral responsibility from a Frankfurt case, the philosopher would conclude *x* is true after the question was framed with *y*. However, it's not that simple. Both priming and framing can occur subconsciously or without the philosopher or the participant knowing. Priming and framing can also occur before the subject even arrives at the thought experiment, which is outside of everyone's control. Even if a person had a bad encounter with someone named "Alexander" and the name "Alexander" is used in the thought experiment, then the participant would probably be more likely to place blame on "Alexander." Or, it can be as unobvious as having an insignificant experience or encounter that subconsciously resurfaces to influence how a person feels about a scenario. Thus, it is impossible to account for all possible priming and framing biases. Even if only a few people are experiencing these effects, it still poses a threat to thought experiments because the philosophers cannot be aware of the occurrences.

A. Instability of Beliefs and How Beliefs Change Over Time:

Another factor that can significantly bias a philosopher's thought experiment is the fact that as humans, our intuitions aren't only shaped by society and situational factors, but our intuitions also naturally evolve over time; they do not remain constant forever. In "Whose Concepts Are They Anyway? The Role of Philosophical Intuition in Empirical Psychology," Alison Gopnik and Eric Schwitzgebel prove that as we age, our intuitions evolve. They claim that we have to

be careful about relying on our intuitions because our intuitions are unstable. They demonstrate this shift in intuitions by showing that children's intuitions and adult's intuitions differ. Thus, an intuition is not a reliable source as a means to understand overall human thought because a person's intuitions are constantly evolving. Because everyone's intuitions develop at different times, it might be hard to know what stage of development a person is in while surveying his or her intuitions. Even if the researcher were to limit his subjects to adults with "developed" intuitions, then he would still not get an accurate picture of intuitions overall because we do have to account for the fact that intuitions do develop and change over time. Although this finding might call into question the value of intuitions more than it condemns the use of thought experiments, the mere fact that philosophers have no way of knowing what stage of development a subject is in, he or she cannot draw a conclusion about a belief in general because it is proven that a person's intuitive response will at some point evolve.

If the philosopher conducting a thought experiment is not careful about properly selecting and documenting his or her participants' age, then this finding can be a biasing factor in the thought experiment results. For example, if a researcher is using Frankfurt's thought experiment to try to gain insight about beliefs regarding moral responsibility, he or she needs to keep in mind that one's beliefs will most likely not be consistent forever. If the researcher surveys too many younger people compared to older people, or vice versa, then the data will be skewed. Because philosophers are not always as careful as scientists in selecting

their survey sample, it's hard to always account for what stage of development a person is in.

Not only do our intuitions change as we age, but our beliefs also change given the amount of time we have to contemplate an action. Antti Kauppinen also notes that the response a person gives in the heat of the moment is usually temporary and does not reflect the person's overall belief about the topic because, after given more time to think and reflect, a person's response will very likely change. He examines Knobe's CEO vignettes thought experiment, which tests intuitions about intentionality in either "harm" or a "help" conditions. Kauppinen's research suggests, according to Weigel, that people find, "that the CEO did not intentionally harm the environment, but then a few days or weeks or months later can and often do go on to attribute intention to action on the basis of moral consideration" (Weigel, 2009, 239). Not only do our responses change over time, but, given more processing time, our beliefs also appear to change. This does not threaten the use of intuitions because intuitions are not concerned with beliefs after deliberation; however, if an experimental philosophers want to get an accurate understanding of human beliefs on a particular topic, then the amount of time a person has to processes the experiment will influence how a person views the situation or scenario presented. Thus, a person's immediate reaction to a Frankfurt experiment and whether or not he chooses to ascribe moral responsibility might indicate one belief, but then over time, that belief can definitely change.

B. Analysis of Thought Experiment Literature

Many philosophers who support experimental philosophy don't necessarily agree with all of the practices used, but, rather, hold that the movement reveals useful information about armchair theorizing and how empirical data can be of use to philosophy. One proponent of this view is Chris Weigel. In her paper entitled "Experimental Philosophy Is Here to Stay," Weigel claims, "experimental philosophy raises important questions about methodology, opening the door on new questions and new ways of looking at old questions" (Weigel, 2009, 221). I think Weigel is right that we need to examine the methodology of experimental philosophy. In this paper I have pointed out many reasons to support the claim that the methodology of thought experiments in experimental philosophy leads to unreliable data. However, Weigel argues that this methodology opens the door to new ways of looking at old questions. I do agree with Weigel that thought experiments have provided a new tool that allows us to see old philosophical questions in a new light. However, my beliefs diverge from Weigel's because I do not find that thought experiments have succeeded in accurately answering the questions they set out to examine in regards to moral responsibility and ethics. Just because we might not get an accurate answer from these experiments does not mean that we aren't learning new things.

Experimental philosophers might be charged with sometimes drawing the wrong conclusions from the questions they ask, but we can still learn a lot about various reasoning traits and human perception. These perspectives might not have been of concern to philosophy earlier, but they can shed light on topics that philosophers have been very concerned with, such as knowledge and moral

responsibility. If the philosopher sets out to gain an understanding of when people generally attribute moral responsibility and what he finds instead is a new way of looking at an old question or realizes that we should amend an accepted principle, then this does not make the project of learning about the contours of attributions of moral responsibility successful. Rather, we have learned something new or challenged something that has been widely accepted. The overall goal was not achieved, but the new thing discovered can turn out to be incredibly significant and can enhance or alter a philosopher's perspective on a given topic. In other words, the conclusion does not follow from the hypothesis, but that does not necessarily render the conclusion useless or unproductive.

Another point that I do not agree with Weigel on is her idea that we can try to eliminate cognitive biases. She writes, "we are in a much better position to answer whether free will is compatible with determinism if we know what our cognitive biases are and strive to eliminate them first" (Weigel, 2009, 233). It's not that striving to eliminate our cognitive biases will change whether or not we find free will and determinism compatible, but, rather, we will be able to get a more accurate understanding of a person's beliefs if we are able to eliminate his biases. In theory, this is a great idea, but, however, it is not practically possible. It is very easy to say that once we're aware of our biases we can eliminate them. There are two reasons as to why eliminating our biases is not possible: 1) we are not conscious of many of our biases and 2) many biasing factors are out of our control. It is impossible to limit or control factors that are embedded in our brains. Many factors that influence how we answer a question have been imposed on us by our society and upbringing,

which are influences that lead to us think and reason in a certain way. We also cannot completely control the things or people that we encounter and experience and those who we continuously learn from, such as teachers and parents. These influences that are out of our control will teach us to reason in a certain way and after being reinforced many times, these ways become natural and second nature to us.

Another one of Wiegle's points that I'd like to draw on is about her view of the surveys that experimental philosophers use. She argues, "… experimental philosophers administer the surveys in a way that minimizes subject fatigue and other performance errors" (Wiegel, 2009, 238). My criticism of this argument is very similar to my criticism of her claim that we can account for our cognitive biases. The two claims parallel each other because researchers cannot alter the survey to accommodate each individual's performance errors. According to the NFC factor, some people are inherently more susceptible to the effects of priming and framing than others. The researcher can try to account for some performance errors by avoiding or considering a bystander effect or using abstract compared to concrete conditions, but not all performance errors can be minimized enough to make the results reliable. We are not always completely aware of the things that can cause performance errors in all individuals.

In an interview she gave in 3:AM Magazine from October 26th, 2012, Weigle once again defends experimental philosophy and explains that ex-phi allows her to explore free will questions that are perennial with concepts and ideas from social psychology. However, I believe that if philosophers want to use psychological

concepts to collect their data, then they should accurately replicate how the psychologists do it. Philosophers can't claim to use psychological methods if they are not following all of the rules of psychological practices. Psychologists follow a strict scientific method that ensures their data is collected accurately. Also, their results have to be able to be replicated to ensure accuracy. Philosophers do not follow and have not been trained in how to follow these guidelines, rendering their "scientific method," or lack thereof, unreliable in the eyes of the psychological world. However, even if philosophers were able to account for the biases inherent in the thought experiments, I cannot say with certainty that they would still be able to accurately answer philosophical questions on topics such as moral responsibility, free will, etc. Although we might not be able to draw complete conclusions on these topics from thought experiments, we can still open the door to new questions on these topics.

Another philosopher who is a strong advocate for the use of intuitions and experimental philosophy is George Bealer. In his article, "Intuition and the Autonomy of Philosophy," Bealer advocates for the use of thought experiments in terms of studying 'physical intuitions:' "traditionally, in a thought experiment one usually elicits a physical intuition (not a rational intuition) about what would happen in a hypothetical situation in which a physical, or natural, laws (whatever they happen to be) are held constant but physical conditions are in various other respects nonactual and often highly idealized" (DePaul & Ramsey, 1998, 207). What Bealer means by "physical intuition" is an account of how people visualize a certain situation or circumstance without actually seeing the object. In other words,

thought experiments are successful in terms of understanding how people think about physical situations rather than understanding what people think about certain principles. For example, Newton's famous thought experiment about a rotating bucket in empty space studies how people intuitively envision the physical properties of the bucket and water. But, according to Bealer, thought experiments are not meant to test how people think about non-physical entities, such as morals. Even though many current experimental philosophers do use thought experiments to try to understand what people think about morals, Bealer- although in favor of the use of intuitions- would most likely not support those claims and consider the findings a "belief" rather than an accurate conclusion based on intuitions.

Bealer wants to distinguish this use of intuition from beliefs: "belief is not a seeming; intuition is" (DePaul & Ramsey, 1998, 208). When thought experiments try to test ideas concerning morals and ethics, they are addressing a belief. Thus, thought experiments cannot accurately address these topics. Even though Bealer is in favor of "thought experiments" he would, most likely, not back Frankfurt's thought experiment because Frankfurt is testing moral responsibility, which is a subject that Bealer would consider a "belief" and not a 'physical intuition.' Therefore, even an advocate of the use of intuitions, Bealer, would not support the data obtained by Frankfurt's thought experiment and other similar results.

C. Conclusion

With the growing awareness of experimental philosophy, philosophers conducting thought experiments need to be leery of how they are implementing

their thought experiments. There are many factors that cause biases in their studies, most of which are unavoidable, including how the scenario is worded, differences in human brains, social norms, and individual experiences. Not only does each bias account for a substantial inadequacy individually, but also when we consider all of the biases that are possibly influencing each thought experiment together, it becomes clear that no thought experiment can possibly produce unbiased responses. Even if the philosopher took the time to analyze each and every participant to check for individualizing characteristics that could affect his or her response, he could still not factor these characteristics in because many of the biasing factors are subconscious. Also, by surveying people for the abovementioned biasing factors, one could actually induce further biases.

Even when a subject is asked to consider his personal characteristics, certain stereotypes can be activated, which can make a big difference in how the subject responds to a thought experiment. For example, a study done by Harvard University Department of Psychology proved that "Asian-American women performed better on a mathematics test when their ethnic identity was activated, but worse when their gender identity was activated, compared to a control group who had neither identity activated" (Shih, Pittinsky, & Ambady). Because it is generally believed that Asians possess better quantitative skills, the Asian-American women performed better on a math test when they were reminded of their ethnicity compared to when they were reminded of

their gender because it is a cultural stereotype that females have inferior quantitative skills compared to males. Thus, by acknowledging certain identities

right before a thought experiment, the identity will be more prevalent in the person's mind and can emerge to make a difference in answers. This further reinforces the claim that biases in thought experiments are unavoidable because the philosopher cannot know if an individual has a specific identity that has been "activated." Until philosophers can agree on and establish some sort of "scientific method" for thought experiments in order to avoid most of these biasing factors, these experiments cannot be considered precise tools to account for philosophical conclusions.

Despite having pointed out all of the abovementioned biases inherent to thought experiments, I am not arguing to put an end to experimental philosophy. I think that it is important for experimental philosophers to be aware of these biases and the fact that there is no way to account for them. Even though we cannot completely conclude what people think about moral responsibility from a thought experiment similar to that conducted by Harry Frankfurt, it is possible that we can learn something new regarding the topic at question. Although the thought experiment might not accurately answer the question it seeks to examine, by just creating the thought experiment the philosopher can create a new perspective on the topic. We cannot let the perfect be the enemy of the good; we know that thought experiments are far from perfect, but if we are learning new things from them, then they can prove to be useful for philosophy and we should continue building on experimental philosophy.

One experimental philosopher who made a big break in epistemology by designing a thought experiment is Edmund Gettier. For hundreds of years people

had accepted Plato's claim that knowledge is a justified true belief. It wasn't until the 1960's when Gettier published a thought experiment in his paper "Is Justified True Belief Knowledge?" that details a breakdown in information. It shows that what one thinks is a "justified true belief" might not necessarily be a clear case of knowledge. Even though Gettier's thought experiment is susceptible to the same biasing factors that Frankfurt's thought experiment and all other thought experiments are predisposed to, he managed to challenge one of the most widely accepted beliefs in philosophy. Similarly, even though Frankfurt could not sufficiently conclude what people think about moral responsibility from his thought experiment, he was able to suggest an amendment to what had been a widely accepted principle, the principle of alternate possibilities.

Overall, I do believe that philosophers need to be skeptical when using intuitions obtained by thought experiments as data for drawing philosophical conclusions. However, my skeptical mindset regarding these thought experiments does not serve to completely undermine my belief in the accomplishments of experimental philosophy. Due to the fact that potential biases are unavoidable in almost all situations, we should not let these biases prevent experimental philosophers from seeking to gain a new perspective on a philosophical question. After examining the same questions for hundreds and hundreds of years, philosophers need to take new approaches if they want to gain new insights; thus, I commend experimental philosophers for seeking to expand the realm of philosophy into that which is concerned with empirical data. Although we are realizing that not all empirical data can be trusted for philosophical conclusions, by taking the time to

examine certain questions, such as those about knowledge, moral responsibility, free will, etc. that have been discussed in the philosophical world practically forever, we can make gains on the topics that would otherwise remain stagnant. Yes, thought experiments have their flaws, but experimental philosophers should continue to pursue the questions that they pose.

Works Cited:

Bealer, George. "Intuition and the Autonomy of Philosophy." Trans. Array *Rethinking Intuition*. Michael DePaul and William Ramsey. Lanham: Rowman & Littlefield Publishers, Inc., 1998. 201-239. Print.

Buckwalter, Wesley, and Stephen Stich. "Gender and Philosophical intuition." (2010): 1- 39. Print.

Fischer, John. "Philosophical Review." *Philosophical Review*. 119.3 (2010): 315-336. Print.

Frankfurt, Harry. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy*. 66.23 (1969): 829-839. Print.

Gopnik, Alison, and Schwitzgebel Eric. "Whose Concepts are they, Anyway?: The Role of Philosophical Intuition in Empirical Psychology." *Rethinking Intuition*. Ed. Michael DePaul and Ed. William Ramsey. Rowman & Littlefield Publishers, Inc., 1998. 75-91. Print.

Gendler, Tamar. "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium." Midwest Studies in Philosophy. (2007): 68-89. Print.

Gettier, Edmund. "Is Justified True Belief Knowledge?." Analysis. 23. (1963): 121-123. Print.

Marchall, Richard. "X-phi is here to stay." *3:AM Magazine*. 26 2012. Print.

Nichols, Shaun, and Joshua Knobe. "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions." NOUS. 41.4 (2007): 663-685. Print.

Shih, Margaret, Pittinsky Todd, and Ambady Nalini. "Stereotype Susceptibility: Shifts in Quantitative Performance from Socio-Cultural Identification." Web. <http://icos.groups.si.umich.edu/shihpaper.html>.

Weigel, Christine. "Experimental Philosophy is Here to Stay." *Analyse & Kritik*. 02. (2009): 221-242. Print.

Whitley, Bernard, and Mary Kite. *The Psychology of Prejudice and Discrimination*. 2nd ed. Belmont: Wadsworth, 2010. 178-184. Print.