5-2021

Weight of Words: Moral Responsibility and Freedom of Speech

Sihan Feng

Weight of Words: Moral Responsibility and Freedom of Speech

A thesis submitted in partial fulfillment of the requirement
for the degree of Bachelor of Arts in Philosophy from
The College of William and Mary

by

Martin Sihan Feng

Accepted for _____Honors_____

_____*Philip Swenson*_____

Philip Swenson, Ph. D., Director

_____*Jennifer Bestman*_____

Jennifer Bestman, Ph.D.

_____*Tim Costelloe*_____

Timothy Costelloe, Ph.D.

Williamsburg, VA
May 11, 2021

WEIGHT OF WORDS: MORAL RESPONSIBILITY AND FREEDOM OF SPEECH

Martin Sihan Feng

April 2021

I. Introduction

In this thesis, I will propose a moral responsibility framework termed "the Anticipation Model," which argues that for an agent to be held morally blameworthy for any act, two necessary conditions are required. First, they can freely choose not to perform the action, and second, the committed act either violates their normative judgment at the time of action or violates the agent's general moral beliefs. Based on the above moral framework, I will subsequently defend freedom of speech through arguing that a positive moral responsibility judgment for speech is seldom justified. If, under rare circumstances, speech responsibility can be determined, people still ought to be skeptical about the amount of blameworthiness that can be rightfully attributed to the speaker.

II. The Anticipation Model of Moral Responsibility

We have the intuition that an agent is not responsible, especially not blameworthy, for their action if they do not think what they are doing is wrong. Such a principle, despite being appealing, provides a rather crude normative guidance. It includes much of the worst behaviors, such as murder without regret, that most of us reprehend. This reveals an intuitive conflict of our daily moral judgement – on the one hand, it seems overly demanding to hold one responsible for acts that they judge to be the morally best; on the other hand, individual's moral psychology varies considerably to the extent that certain moral judgment is simply deemed as being intrinsically detestable.

The first half of this article aims to solve this tension by addressing the following question: under what conditions can one be reasonably held responsible? Although many agree that understanding the results of one's action is a necessary condition for responsibility – because envisioning the possible outcomes is essential to decision making – more questions naturally

arise. What knowledge, both descriptive and normative, should the agent be expected to have about such outcomes? Humans are not omniscient, and our predictions of the future are seldom accurate let alone perfect. Much of people's anticipation is a point-to-point deduction. For example, I would anticipate that if I throw a baseball to a nearby window, I will shatter it. This simple method renders a quick comparison between choices, which allows decisions to be made rationally. But our prediction should be taken as a simple sketch of the complicated reality, if not entirely detached from the truth. As we make anticipations about more complex events, it becomes more difficult to foresee the outcomes. In addition, the causal relationship between one's action and the outcome may not be a linear one. The consequences resulted from one's action can often be contributed by multiple causes. Me throwing a baseball, for instance, does not necessitate the shattering of the window, whose happening is therefore not entirely within my control. For such an event to occur, many conditions are required: a clear path between me and the window, a window made of glass or other fragile materials, a well-made baseball that is sturdy enough to break the window, and so on. As a result, people should acknowledge that human perception and prediction of the world is a mere attenuation of truth at best. It is natural to attribute the shattering of the window to my act, as if it is the only cause, but it must be recognized that me alone, along with any other agents in the world, does not have full control nor complete understanding of the future even though some future events are causally related to my behavior. The uncertainty between the attitude, action, and consequence appears to violate the notion that one can only be held responsible for what one has control of. If I cannot fully control the outcome of my action – whether the window will break or not – how can I be responsible for breaking it? This question reveals a responsibility issue known as the resultant moral luck. The term is developed by Thomas Nagel, who suggests that "Actual results influence culpability or

esteem in a large class of unquestionably ethical cases ranging from negligence through political choice." [1]

I contend that to minimize the issue of moral luck and to locate the origin of responsibility, one should adopt the Anticipation Model of Moral Responsibility, whose definition goes as follows:

> **The Anticipation Model of Moral Responsibility**: An agent X is blameworthy for their action A if X could have refrained from doing A freely and 1). X makes a positive moral judgment about A and believes that choosing A is morally wrong at the time of action, or 2). choosing A goes against X's general moral belief when X makes no moral judgment about A at the time of action.

According to the anticipation model, responsibility stems from two possible scenarios: first, the agent commits an action freely, knowing what they are about to do is wrong. Second, the agent commits an action freely without judging whether the action is wrong or not, but this choice violates their moral beliefs that they normally have. For example, assuming there are no constraints on my action, I am blameworthy for killing my friend Priscilla if I think that killing her is impermissible right before I initiate my act. I can also be blameworthy if I do not have an opinion about killing while it is happening, but I generally hold the belief that killing is wrong (a.k.a. the general moral belief).

III. Defending the first part of the Anticipation Model

The first part of the anticipation model follows the principle of "alternate possibilities" (PAP), which states that an individual is only morally responsible for an action if he could have done otherwise. Though many takes this principle to be a fact, Harry Frankfurt attempts to deny it through a renowned thought experiment, where he presents a case that seemingly has no alternatives yet still allows one to be held responsible. Similar examples offered later by other

[1] Thomas Nagel, *Moral Luck*, 24-38.

philosophers, such as John Martin Fischer,[2] are known as Frankfurt-style or Frankfurtian cases. Surprisingly, the nature of "alternate possibilities" is often neglected, thereby generating various confusions. I will, thus, defend PAP by analyzing how two kinds of alternatives – the alternate option and the alternate will – are required for responsibility. I will consequently show how Frankfurtian cases only deny the alternate option, hence fail to undermine PAP.

A typical Frankfurt case is provided below:

> **Chip**: Jones is about to cast a vote for A. Unbeknownst to Jones, a manipulative chip is secretly inserted in his brain. If Jones shows a prior sign of deciding to vote anyone other than A, this chip will effectively change Jones' mind and makes him vote for A instead. However, this chip never needs to be activated because Jones votes for A out of his own reasons.[3]

Frankfurt thinks that Jones is clearly responsible for his action, though he could not have done otherwise, therefore rejecting the PAP. However, the phrase "could not have done otherwise" does not accurately capture the meaning of "alternate possibilities." Commonly, being able to do otherwise emphasizes the option and demands a person to make a difference in the real world. On the other hand, having alternate possibilities is more nuanced. Besides being able to do otherwise, one can possess alternate possibilities by having mental alternatives that enable them to will differently. In my opinion, there are two kinds of morally relevant alternatives, as briefed above, the alternate option and the alternate will. The alternate option represents the objective choices available to the agent, which are physical actions that can be achieved if one desires to. Denial of the alternate option can limit one's freedom and in extreme circumstances lead to coercion. In contrast, the alternate will strictly occurs within one's mind. This is the different favorability based on the same ideas or beliefs that ultimately contribute to the choice.

[2] John Martin Fischer, *Semicompatibilism*, 40-41
[3] Harry G. Frankfurt, *Alternate Possibilities and Moral Responsibility*, 831

It then becomes clear that, although Jones is not capable of actually making a change (voting a different person), he is able to follow a different mental path even it leads to the same result. Fisher has argued that this difference is sufficient for holding Jones accountable.[4] This is referred to as the "fine-grained" approach. Although Jones will inevitably vote for A in any circumstance, he has the flicker of freedom between voting A out of himself or out of the chip's influence. Had Jones omitted to vote for A out of his own, he would not been responsible for this action.

I argue that the alternate will is required for one to be responsible. Imagine the case followed:

> **Free Murderer**: Elizabeth is a bloodthirsty murderer. She enjoys others' pain and fantasizes killing random people she finds on the street. Surprisingly, Elizabeth has a normal childhood with a common experience. On a Friday afternoon, she murdered Martin in a brutal way but was quickly caught by the local police.

Police charges Elizabeth with the first-degree murder, and specifically tries to hold her responsible for Martin's death. Elizabeth admits the crime and goes to jail, believing that what she does is wrong. In accord with our intuition, the anticipation model judges Elizabeth to be blameworthy for murder.

> **Mad Scientist**: Learning about Elizabeth's case, a mad scientist, Joseph, copied the entire personality of Elizabeth and stored the information into a thumb drive. In a sense, Elizabeth is "resurrected." Joseph wiped out all the data that happens right after she murders Martin and inserted the thumb drive into a super simulator that recreates the experience of the murder scene. Inside the simulator, Elizabeth, having the same beliefs and reasons, chooses to spare Martin, which leads her to a different future without going to the jail.

The Mad Scientist case clearly demonstrates that Elizabeth has the alternate will since she is able to form a different decision with the past and laws of physics being the same. Here, the only thing that has altered is Elizabeth's will. However, consider now:

[4] John Martin Fischer, *The Metaphysics of Free Will* (Oxford: Blackwell, 1994), p134

> **Deterministic Murderer**: In a parallel but deterministic universe, the same Elizabeth from the Indeterministic Murderer case, again, murders Martin in the same way and is caught. She admits what she did but claims that punishing her would be unjust. "Had anyone had my genes, my experience, my upbringing, everything I had, this person won't be able to make a difference but to murder Martin."

The Mad Scientist Joseph from the deterministic universe also copies the personality of Elizabeth and connects it to a simulator. He also makes the copied Elizabeth re-experience the same Friday afternoon, but every time Elizabeth chooses murder. Joseph, to his surprise, found that no matter how much memory he deletes from Elizabeth, so long he runs the same experience for her, Elizabeth still definitively murders Martin on the exact same Friday afternoon. The interaction between her circumstances and her brain composition not only necessitates her decision on murdering Martin but also all of Elizabeth's choices throughout her entire life. Her fate is determined the moment she was born, and she never enjoys the alternate will. Consequently, she is not responsible for murdering Martin as all of her decisions are ultimately pushed by her upbrings, environment, laws of physics and other factors beyond her own control.

So far, I have demonstrated the intuitive moral judgement regarding whether the agent enjoys the alternate will, but advocates of Frankfurt might argue that the intuition from the Deterministic Murderer case is misleading. Although the only difference between the Deterministic Murderer case and the Chip case is the flicker of freedom, Frankfurtians can argue that such freedom is not significant enough for responsibility.

One of the responses to the above argument is the "dilemma defense", which states that it is impossible for Frankfurtian cases to deny both the alternate will and a prior inference. If the prior sign does not determine what Jones will do, the interferer, such as a manipulative chip, can only intervene when the decision has already been made, making the flicker of freedom a robust decision. Alternatively, if the prior sign does accurately predict Jones' decision, then Jones is not responsible. However, one might claim that the dilemma defense relies heavily on the notion of

prior sign, whose existence itself is questionable. I think it matters not whether the agent is actually capable of exhibiting a sign or a flicker of freedom, but rather whether the interferer is necessary, which is determined by the decision that one would have made. Consider the following case:

> **God**: God carefully monitors the interaction between the simulator and the thumb drive. As an omniscient being, God simply knows the result of Elizabeth's decision, even if there is no prior sign at all. God instructs this information to Joseph, who rewinds the time back to the same Friday continuously. Joseph only connects the thumb drive to the simulator if God tells him that Elizabeth decides to murder Martin. If God tells the otherwise, Elizabeth won't exist at all .

With the aid of God, Joseph does not rely on any prior sign to decide whether the thumb drive should be connected. He simply knows it as God does. Since Joseph does not intervene with the decision of Elizabeth – he simply refrains Elizabeth from having the opportunity to make a decision – Elizabeth won't be able to demonstrate even the slightest amount of freedom.

I contend that Elizabeth is responsible in the God case, but she still can satisfy PAP by retaining the alternate will. A perfect being like God may predict that whether Elizabeth is going to kill with certainty, yet had Elizabeth preferred the alternatives instead (suggesting that she enjoys the alternate will), God would have made a different prediction. In other words, suppose Elizabeth would have chosen not to kill, then some form of intervention is required, whose presence of itself grants the robustness for the alternate will. If the presence of an interferer is irrelevant, it entails that Elizabeth never has the alternate will and therefore is not responsible.

By assuming the omniscience is compatible with free will, I show that one can be perfectly predictable but still enjoys her alternate will and thus be held accountable. Elizabeth may never exert her alternate will or the flicker of freedom in real world, but she can still possess it to satisfy the PAP. Even Elizabeth is unconscious and stored in a thumb drive at the moment, what she would have done is an unchangeable fact. This ability of leaning toward a different decision itself upholds PAP.

Some objections to my argument target the Deterministic Murderer case. Many acknowledge that Elizabeth is not able to refrain from killing Martin without alternate will, but still believe that she is blameworthy. With or without the alternate will, Elizabeth still deliberates among the same reasons and values prior to the killing. Had the indeterministic Elizabeth decided the same as the deterministic Elizabeth, the thought process and the person is identical, which grounds her responsibility. I name this objection the Conceived Alternative objection.

The Conceived Alternative objection is appealing, but it only offers practical benefits rather than providing a justification. While Elizabeth conceives exactly the same alternatives with or without alternate will, she only has control over her actions with it, which irreducibly determines the future of her own. On the contrary, the Elizabeth without alternate will is not able to enjoy such control, albeit sharing the same ideas, reasons, and circumstances. She is pushed by her past experience and laws of physics, which definitively fixates her decisions. She has no power over what kind of decisions she makes, but to accept and act on the decisions passively.

In conclusion, I have clarified the ambiguity of the PAP, which does not clearly specify the meaning of alternate possibilities. Specifically, I separate the alternate possibilities into two kinds: the alternate option and the alternate will and contend that they are both morally relevant and necessary for responsibility. Frankfurt-style examples, I argue, fail to coherently make the absence of alternate possibilities co-exist with free will. Additionally, I demonstrate that it is irrelevant if the agent in fact utilizes his freedom or has a chance to demonstrate his freedom. My response to the Perceived Alternatives objections indicates that the alternate will is vital for our responsibility since it grants us the ultimate control over our actions. I conclude that Frankfurtian cases do not undermine the principle of alternate possibilities and moral responsibility does require one to freely choose not to do action.

IV. Defending the second part of the Anticipation Model

As mentioned above, due to the limitation of our cognitive capacity, what goes into our decision reasoning might not be reflective to what is happening in the real world. Nevertheless, I contend that the strict adherence between our conception of the world is not at all required for responsibility. Consider Matt, who firmly believes that sugar is fatal for humans, as an example. He adds excessive sugar to his neighbor Isabella's coffee under the impression that it will kill her. Matt is partially out of touch with reality, yet many, including myself, still consider him to be blameworthy. Had Matt been right, he would have successfully murdered Isabella. In this particular case, Matt bears a clear will and actualizes it through "poisoning" – putting sugar in his neighbor's coffee. While whether Matt deserves punishment in a real-world scenario is up to debate, but this case still demonstrates that it is not inconceivable to ascribe blameworthiness to someone with incorrect understanding of the reality.

The argument can be further advanced so that objective reality itself is not even required for responsibility and one's degree of responsibility is entirely dictated by his or her own anticipations. Consider an alternative case of Matt:

> Matt is completely illusionary and hospitalized for paralysis. He carefully deliberates between options of killing his neighbor Isabella. He chooses sugar since he thinks this would causes most pain to her, but little did Matt know, Isabella, his home, and the world he sees are all fictional and only exist in his mind.

In the second case, Matt is completely detached from reality, however, he still appears to be culpable for poisoning though he only does it in his mind. The delusional Matt has more detailed expectation of sugar poisoning – instead of simply murdering Isabella, he believes that sugar would inflict severe pain on her. While Matt is further away from the reality, he is, I argue, no less blameworthy compared with Matt in the first case.

Since the anticipation model does not rely on the actual outcome to decide one's responsibility, it resolves the issue of resultant moral luck. Consider a second scenario of baseball throwing. Jack made the same effort calculating, aiming, and throwing the ball as I did, but the window did not shatter. In fact, in the second scenario the window was not affected at all. By sheer luck, the exact same action leads to two distinct outcomes: I break the window and Jack does not. Suppose that both me and Jack think that breaking others' window just for fun is wrong, yet at the time of action, we still throw the ball. The resultant moral luck challenge suggests that I am responsible for breaking the window while Jack is not responsible. This is counter-intuitive because the difference in result is beyond our control. In response, the anticipation model holds that since we both regard our action to be wrong at the time of action and we perform the same action, Jack and I are equally responsible.

The resultant moral luck objection might pressure on such a reasoning by appealing to the subject of responsibility. In other words, how can me and Jack be equally responsible when I am responsible for breaking the glass and Jack appears to be responsible for breaking nothing? Being a good citizen, I also feel obliged to pay for the broken glass and apologize for my behavior, while Jack certainly has no reason to do the same. Similarly, even though the delusionary Matt is not likeable, it seems harsh to blame Matt for murdering in the same way we might blame an actual killer. Afterall, Matt does not actually cause any harm.

To respond, the anticipation model shifts the subject of responsibility from the real-world consequence to the subjective intention of the agent. The ground of my responsibility lies on my anticipation and moral judgement of my behavior rather than what actually happens. Regardless of whether determinism is true or not, it is near impossible to predict what an action is capable of leading to. This asymmetry between our perception and objective truth is what gives rise to the

resultant luck problem. Therefore, instead of asking what the actual consequence is, the anticipation model asks what the consequence would be had the agent been right, which strictly puts the basis of responsibility within the control of the person.

The anticipation model directly challenges the sane deep-self view proposed by Susan Wolf, who, siding with Frankfurt, Watson, and Taylor, believes an agent is responsible for their action if such an action is a result of their will and such a will is within the control of their deeper selves.[5] In simpler terms, one is responsible for doing an action if he both wants to do the action and he wants to want to do the action. Nevertheless, Wolf is not satisfied with the plain deeper-self view as the origin of the deeper-self can still be beyond one's control. Consider the case of Jojo:

> JoJo is the favorite son of Jo the First, an evil and sadistic dictator of a small, undeveloped country. Because of his father's special feelings for the boy, JoJo is given a special education and is allowed to accompany his father and observe his daily routine. In light of this treatment, it is not surprising that little JoJo takes his father as a role model and develops values very much like Dad's. As an adult, he does many of the same sorts of things his father did, including sending people to prison or to death or to torture chambers on the basis of whim. He is not coerced to do these things, he acts according to his own desires. Moreover, these are desires he wholly wants to have. When he steps back and asks, "Do I really want to be this sort of person?" his answer is resoundingly "Yes," for this way of life expresses a crazy sort of power that forms part of his deepest ideal.[6]

Wolf believes that Jojo is not responsible for all the cruel things he does – "In light of JoJo's heritage and upbringing-both of which he was powerless to control, it is dubious at best that he should be regarded as responsible for what he does" (379-380). However, under the plain deep-self view, Jojo satisfies both conditions and is therefore responsible. Interestingly, Wolf contends that Jojo is not responsible since he does not have a sane deep self, not because his deep self is beyond his control. To achieve the latter requires one to have the ability of literal self-creation,

[5] Susan Wolf, *Sanity and the Metaphysics of Responsibility*, 376
[6] Ibid., 379

which is "not just empirically, but logically impossible."[7] Therefore, Wolf puts forward the sane deep-self view, a special kind of plain deep-self view that is supplemented with sanity:

> **The Sane Deep-self View**: "Recognizing that in order to be responsible lor our actions, we have to be responsible for ourselves, the sane deep-self view analyzes what is necessary in order to be responsible for ourselves as (1) the ability to evaluate ourselves sensibly and accurately, and (2) the ability to transform ourselves insofar as our evaluation tells us to do so. We may understand the exercise of these abilities as a process whereby we take responsibility for the selves that we are but did not ultimately create."[8]

More precisely, Wolf believes that one is sane when they can 1). understand what is morally correct or good and 2). modify their characters and actions in light of their moral knowledge.

Nevertheless, the sane deep-self view introduces more issues and is far from perfect. Proponents of Wolf only push the origin of responsibility further to the concept of sanity. Yet to be sane depends on contested moral knowledge that is subject to change. Racial segregation was deemed permissible decades ago, but it is far-fetched to claim that most people back then were insane. Also, as shown by the illusionary Matt case, being insane or out of reality does not necessarily negate one's moral responsibility.

In comparison, one advantage of the anticipation model is its action specificity. Wolf's definition of sanity aims to describe the overall mental status of a person. Agents who are sane can still be ignorant on certain facts, which directly impacts their normative judgement. Alternatively, agents who are insane still retain the possibility, albeit slim, of being responsible. Being insane, like Matt, does not necessarily deprive one's moral value or ability to foresee the potential future of its action, though these events are likely to happen. This also explains the intuition that delusionary or insane people are not responsible for their actions. Insanity also can rid one's control over their behavior altogether, making it difficult to discern responsibility.

[7] Susan Wolf, *Sanity and the Metaphysics of Responsibility*, 380
[8] Ibid, 381

In case of Jojo, thus, the anticipation model contends that he is not blameworthy since not only does Jojo never believe that his action, during the course of each individual action, is morally wrong, but also none of his cruelty goes against his general moral beliefs. For Jojo, not killing would be the "wrong" act. I argue that it is incredibly difficult, if not impossible, to demand Jojo not to do what his father does and therefore he cannot be reasonably held responsible.

However, there is still some merit to Wolf's reasoning. The anticipation model agrees with Wolf that moral responsibility does not require literal self-creation. How one acquires his moral belief or how he judges what is the best normatively is not always within one's control, but it should not interfere with his responsibility. Resultantly, the anticipation model does not require a complete, thorough moral reasoning. It simply asks whether the agent holds a normative belief about the action he is about to conduct, but in what ways does the agent reach to such a conclusion matter not. It appears to me, in the very least, better moral reasoning is always possible due to the various constraints, such as time, prior initiating the action. It then remains dubious whether there really exists objective moral knowledge that can be reasonably expected from everyone. One can only consciously take in part of the information from the world, making it incredibly difficult for one to even be aware of all the possible choices. Consequently, people rarely, if not never, act on "complete" moral reasoning, but even with crude, incomplete, or even insane thought processes, we are capable of forming a belief about whether the act is wrong. The anticipation model takes such a belief to be a possible basis of responsibility without any emphasis on its history. Suppose Bill holds the belief that stabbing the voodoo doll will kill his neighbor and killing his neighbor is the wrong thing to do. Let us also assume that these ideas are forcibly put into his mind by a chip in his brain. What Bill has is, in essence, an insane reasoning

process. Bill, however, still retains the free will of choosing between stabbing and not stabbing. If we really stand in Bill's shoes, stabbing the voodoo doll would supposedly kill his neighbor and be the morally wrong action. In short, even though Bill may not even understand how he gets these ideas, so long Bill believes in them and commits an act that he thinks is wrong, he would still be culpable for stabbing the doll. I argue that, therefore, actions that violate moral judgment resulted from poorly formulated moral reasoning can be taken as the ground of moral responsibility.

The anticipation model also provides an explanation on why we do not hold infants and animals responsible for their actions, even though modern scientific discoveries suggest that certain kinds of animals are capable of rational planning. Monkeys and apes are even able to engage in sophisticated maneuvers. Yet there is no evidence that animals are the same kind of moral agents like human, meaning that there is no evidence for them to form any normative judgement. This is, however, largely descriptively dependent and is likely to be affected by our ever-increasing exploration of animal minds. In addition, it is still unclear if animals enjoy free will and the access to alternatives. A popular view treats animals as organic machines who have no control over their behaviors and are only driven by instincts. If animals are denied with free will and therefore alternatives, they certainly should not be held accountable for their actions. However, this instinct-driven hypothesis is facing greater challenges in the modern era, and evidence suggests that certain animals might be able to make rational choices like we do. If this view were true, I believe it is possible for animals to be morally responsible if they truly are capable of forming a similar kind of normative judgement as humans do. However, our current understanding suggests that though the ability of predicting the future is shared by multiple species, the power of empathizing with other individuals and judging what is morally the best

does appear to be unique to human. Moreover, even if animals or infants were moral agents, it is quite possible that they lack so much knowledge, such as empathy toward other species, that the positive moral judgments they have are largely different from the ones humans usually have. It is difficult for us to empathize with species other than our own, and we can reasonably infer that other animal species would also have a hard time understanding how their actions would affect other species' feelings. It is unlikely that animals would know species other than themselves are sentient. Without true empathy, animals can hardly know what their actions do to others, not to mention if they can comprehend whether their actions cause pain or suffering. It is, thus, likely that animals, such as tigers, think that hunting prey or humans are morally permissible. Therefore, animals are not likely to be blameworthy and even if they are, their responsibility is restricted in matters within their own herds.

I would like to emphasize the difference between factual ignorance and moral ignorance. Factual ignorance is related to events that are descriptive while moral ignorance involves normative judgment. For example, if Rob does not know that stabbing Beth can kill her, Rob is factually ignorant. Alternatively, if Rob does know that stabbing Beth can be fatal but deems such an action to be good or morally permissible, Rob is said to be morally ignorant. Many, based on their intuition, have an asymmetric feeling towards Rob. One is less likely to condemn the factually ignorant Rob but more likely to hold the morally ignorant Rob responsible. Imagine both Robs are present in a legal court, the factually ignorant Rob says: "Had I known stabbing will kill my wife, I would never have done that." On the other hand, the morally ignorant Rob says:" I did understand what it would lead to when I stabbed her, but I do not regret my action since I believe that she deserved it and I did nothing wrong." Our intuitive reaction is that the

latter Rob is not only condemnable but also loathsome. It appears that Rob's lack of regret furthers his blameworthiness.

Nevertheless, the anticipation model does not distinguish factual ignorance from moral ignorance in a morally relevant sense, meaning that neither the factually ignorant Rob nor the morally ignorant Rob is blameworthy. While this statement is seemingly appalling, I argue that under closer scrutiny, the anticipation model is still in accord with our intuition. Suppose Rob is deliberating between whether if he should kill Beth, he concludes that it is for the best if he does so. Demanding the morally ignorant Rob not to kill is asking him to contradict his own reasoning. In essence, from Rob's perspective, following such a request forces him to perform an act that he judges to be wrong, which is a choice that cannot be reasonably expected.

Opponents of the above reasoning are likely to set a normative guidance that is not subjective to one's personal moral judgement to avoid the issue of inconsistency. If the ground for right or wrong is independent from one's subjective judgement, we are able to differentially hold the morally ignorant Rob accountable but not the factually ignorant Rob. In plain terms, we might decide that, for instance, it is wrong to intentionally harm others with no good reason and being morally ignorant does not count as one. People in favor of this view also emphasize that such a system allows moral responsibility to be attributed for an educational purpose. Through blaming the morally ignorant Rob, he, along with others who have similar moral judgment, is taught to correct his moral ignorance.

However, we have few reasons to believe in the existence of the objectively correct normative guidelines. Even if these doctrines do exist, it remains unclear whether the society is able to discover them. The idea of moral ignorance itself presupposes a "correct" kind of moral understanding that are often defined by the majority, which lacks sufficient ground.

V. Comparing Rosen, Fitzpatrick, and the Anticipation Model

I want to distinguish the Anticipation Model from Rosen's responsibility theory, which centers around what he calls "True Akrasia," or "genuine, full strength akrasia." For one to conduct an act that is genuinely akratic, Rosen thinks that "He would have to know the pertinent facts about his contemplated act. He would have to know that it was wrong. And he would have to know that in the circumstances, all things considered, he should not do it. He would then have to act despite this knowledge."[9] In short, one is culpable for an action if and only if one knows such an action is wrong but still does it despite the knowledge. Rosen believes that if akrasia is the only root for responsibility, it follows that we, due to our lack of access to the complete casual history of one's action, should adopt a skeptical position about moral responsibility – "His position is that be this as it may, confident positive judgments of blameworthiness are never justified."[10]

According to Rosen, there are two possible impostors that can obscure our judgment on identifying genuine akrasia – ordinary moral weakness of will and failure to access one's normative value. The real-world limitations along with epistemic opacity make the imposters indistinguishable from true akrasia since their difference is purely internal and private to the agent. Rosen thinks that true akratic judgment has a clear preference of one action over the other, which remains unchanged before and after the agent's action. On the contrary, plain moral weakling favors no particular choice, but might lead to the same action as the true akrasia does. Since the only ground for responsibility is genuine akratic acts and not moral weaklings, it would be unreasonable for us to confidently judge one's blameworthiness.

[9] Gideon Rosen, *Skepticism about Moral Responsibility*, 307
[10] Ibid., 295

The skeptical position itself, however, relies on epistemic obstacles that are reasoned through Rosen's personal self-reflection. "When I consider my own case and ask whether some weakish act of mine amounts to genuine akrasia as opposed to ordinary moral weakness, I have no trouble identifying tolerably clear cases of the latter; but I confess that I cannot identify clear examples of the former with any confidence."[11] This is refuted by philosopher Fitzpatrick who argues that there is a feasible way of differentiating "impostors" from clear-eyed akrasia. To do so, Fitzpatrick offers a counter reflection:

> "…I often know such things as that I really shouldn't be digging into a heaping bowl of full-fat Belgian Chocolate ice cream given my cholesterol levels and the fact that I've had two helpings already this week. And I know this perfectly well even at the moment I am doing it, as it is transparently imprudent according to standards I myself accept, even taking all things—such as my present enjoyment—into account. I'm just not sufficiently motivated by these normative thoughts, instead giving into gustatory temptation. Each instance may be no big deal in itself, given the long-term and statistical nature of the health issues, but I nonetheless know that it is not what I should be doing here and now."[12]

He goes on to suggest that this kind of akratic acts are apparently different from acts out of temporary normative ignorance, where the agent, at the time of acting, self deceives into genuinely believing that, all things considered, the action is the right choice. One evidence of clear-eyed akrasia that Fitzpatrick provides is the presence of shame or guilt. These psychological experiences, according to Fitzpatrick, give us "at least strong prima facie evidence that the act was not done out of normative ignorance, either general or temporary."[13]

Yet it is questionable whether putting any confidence in shame or guilt is really justified. Even though that Fitzpatrick acknowledges that utilizing such evidence is fallible, he does believe that "the presence of guilt and shame at the time of acting can often provide strong

[11] Gideon Rosen, *Skepticism about Moral Responsibility*, 309
[12] William J. Fitzpatrick, *Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge*, 595
[13] Ibid.

evidence of akrasia."[14] This view is mistaken for two reasons. First, while it is quite common that akrasia is accompanied with shame or guilt, such a connection is never guaranteed. Take the above ice-cream case for example, the agent might feel bad for eating the ice-cream at the time, but it is also possible that the agent does not feel anything. The agent can happily admit that he did the wrong thing, but add "strangely, I don't feel anything." More importantly, shame or guilt can be influenced by non-morally relevant factors. Imagine the agent above is deliberating while taking a scoop in. The ice-cream is so good that although he thinks that excessive eating is certainly not the better choice, but the gluttony overrides his shame or guilt. The same logic is also applicable to more severe wrong doings. A hitman can believe that killing is wrong when assassinating his target, but be happy about his choice since it pays so handsomely. An akratic act, therefore, does not necessarily lead to any particular feeling. On the other hand, recognized by Fitzpatrick himself as well, that shame or guilt does not necessarily suggest an akratic act either – "People can feel shame even for things they don't actually judge to be wrong, simply out of psychological conditioning."[15] Suppose the agent above does feel shame at the time of eating the ice-cream, it is conceivable that he believes that there is nothing wrong with eating too much, but thinking about the consequence of the choice, such as social sanction or obesity, can generate negative emotions about his decision. In short, relying on shameful feelings can generate both false positives and false negatives.

Second, from an epistemic and empirical point of view, the timing of shame or guilt is extremely difficult to gauge, but timing is critical for determining an akratic act. One can feel shame prior, during, or after the action. One can also feel shame at the time but is not

[14] William J. Fitzpatrick, *Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge*, 595
[15] Ibid.

consciously aware of his feeling. However, in many cases we do not have timely access to agent's inner feelings and thus are reliant on the agent's self-report. Unfortunately, even though we may not have, or may not need to have, any compelling reason to doubt the agent's sincerity.

It is nonetheless hard for one to accurately recall what exactly he feels at the time of action. When eating, it is hardly known if the above agent is 1. feeling shame and is consciously aware of the shame, 2. is feeling shame but his mind is too occupied by the ice-cream to be consciously aware of the shame, 3. not feeling the shame at all. What makes things even more obscure is that one can easily confuse the timing of shame. A normatively ignorant agent, based on Fitzpatrick's view, should not feel guilt at least at the time of act, while an akratic agent can feel guilt at the time of act. If we were to dig deeper into the thought process behind any action, however, it becomes clear that precisely locating our feeling of any action is unfeasible. Can we really tell the difference between feeling shame while acting and feeling shame right after acting especially when there are no other witnesses present at the scene? Suppose that, as the agent's friend, you discover that the agent has eaten all the Belgian Chocolate ice-cream. You ask the agent what he was thinking when he was eating. The agent says: "I knew I was doing the wrong thing, and I was not proud of my choice." On the surface of it, we might think that we have pretty good evidence that the agent was acting akratically. We lack evidence to doubt his sincerity and he does show genuine regret and shaming. However, you, one who strictly endorses the rationale offered by Fitzpatrick, pursues deeper, "When exactly did you feel bad during the course of your action? Did you feel guilty before you reach to the spoon, while you were heaping into the bowl, or after you taste the ice-cream?" The agent, possibly being confused, asks: "I cannot say for sure. I know that I made a bad choice, and I did not feel good about it, but why do you care?" We might deem this kind of confusion or inability to provide exact time of shame as a

different kind of "opacity of mind." You respond: "If you feel bad right before committing the act or during the act, that likely suggests that you were acting akratically, which is bad. On the other hand, if you only feel bad after eating the ice-cream and only realize the action was bad afterwards, you were probably acting out of normative ignorance and thus not responsible for eating the ice-cream."

Notice that although the agent is clearly sincere, we still have issues determining whether he should be held responsible for overeating. This problem, I believe, arises from the fact that Fitzpatrick accepts that responsibility can only ground on genuine akrasia for the purpose of argument, which requires the agent to be aware of or to believe that he was not acting out of the better judgement at the time of action. In the following section, I will illustrate how the anticipation model can provide ground of responsibility for these actions, specifically, through consulting the agent's general moral belief. Additionally, I will demonstrate that while shame or guilt does not point to akratic act necessarily, they can be utilized as evidence for moral wrongdoings based on the responsibility ground given by the anticipation model.

The core of Rosen's skepticism about moral responsibility is that an ordinary moral weakling manifests in the same way as a true akrasia at least from our human eyes. The anticipation model, on the contrary, precisely rejects this conclusion. By allowing simpler moral conclusions and the general moral beliefs to be seen as the ground of moral responsibility, the anticipation model allows an easier access to the agent's normative judgement of his action. Though possibly oversimplified, ideas such as killing or harming is wrong, can be sufficient for determining his responsibility. This clarifies the ambiguities of using shame as an indicator of moral wrongdoings. As mentioned in the previous section, because of a similar kind of "opacity of mind," guilt or shame can be attributed to lack of access to normative belief at the time – a

phenomenon where the agent later realizes the moral implication of their act and experience guilt or shame afterwards – or to a genuine akratic act. However, the anticipation deems both cases to be responsible. Rosen implies that even if the agent knows that binge eating ice-cream is wrong and the agent is binge eating, but if the agent fails to connect these two ideas and subsequently recognize that binge eating is wrong at the time of action, the agent is acting non-akratically and therefore not responsible.

In addition, noted by Fitzpatrick, the idea that one should not be blamed based on doing what he thinks to be right at the time of action is intuitive, yet still provides substantial challenges to our daily moral practices – "the behavior is certainly bad, but it is in accord with the agent's blameless normative (mis)understanding of things, and so this is just another example of action done out of nonculpable ignorance. Such cases, however, cover much of the worst behavior we are typically concerned to hold people responsible or blameworthy for."[16] What Fitzpatrick is concerned with is agents with non-mainstream normative beliefs, where the agent does not fail to access their stored knowledge, but instead act out of a "blameless normative (mis)understanding of things." Importantly, contrary to what Fitzpatrick contends. the anticipation model does not hold actions that are based on "incorrect" moral understanding responsible.

To explain this difference, we should ask the question, similar to what Fitzpatrick proposes, that whether it is reasonable for us to expect the agent to do otherwise at the moment of action. Let us consider an example provided by Fitzpatrick:

> "Consider Mr. Potter, a powerful businessman who holds false moral views. He takes certain business practices—such as liquidating Bailey's Building and Loan and sticking it to the poor families of Bedford Falls—to be "permissibly aggressive," when in fact they're reprehensibly ruthless." This leads him to do bad things, though he doesn't

[16] William J. Fitzpatrick, *Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge*, 601

> understand that he's acting badly, which means that he's acting out of a certain kind of ignorance. He's fully aware of all the circumstances, but he applies flawed normative principles or weightings and comes up with bad decisions. Is he culpable for his bad actions?" [17]

Fitzpatrick argues that Mr. Potter has a culpable kind of normative ignorance due to three factors:"

1. There were no relevant limitations in his social context or in his capabilities that should have made the necessary broader reflection and information gathering impossible or unreasonably difficult for him.
2. The failure of adequate reflection and information gathering was instead the result of voluntary exercises of vices such as overconfidence, arrogance, dismissiveness, laziness, dogmaticism, incuriosity, self-indulgence, contempt, and so on.
3. He could thus reasonably have been expected to take steps that would have eliminated that ignorance, by refraining from exercising those vices and instead taking advantage of the epistemically relevant opportunities available to him."

The anticipation model deems all three factors to be implausible. The above reasoning confuses between what is reasonable and what is possible. While it is certainly possible for Mr. Potter to act otherwise, it is not reasonable for us to expect him to do so. First, while there are no relevant limitations for Mr. Potter to make broader reflection, there are also no sufficient motivation for him to do so. It is also not clear whether a broader moral reasoning and exposure to more information will change Mr. Potter's mind. It is possible that Mr. Potter still holds the exact same moral view even after he attempts to educate himself morally.

[17] William J. Fitzpatrick, *Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge*, 599

Second, it appears that Fitzpatrick only attribute Mr. Potter's moral (mis)understanding to negative and voluntary exercises (overconfidence, arrogance, dismissiveness, laziness, dogmaticism, incuriosity, self-indulgence, contempt, and so on), and neglects the possibility that Mr. Potter might actually have a genuinely different moral opinion that was carefully reasoned. Moreover, it is open to debate whether being overconfident, arrogant, dismissive, lazy, and so on really is a voluntary choice. Mr. Potter might acquire these traits through factors beyond his control, such as genetic makeup or childhood upbringing. Even if we were to accept that these are voluntary choices made by Mr. Potter, one can argue that these are non-culpable choices. Again, Mr. Potter might genuinely believe that it is for the best that he is overconfident, arrogant, dismissive, lazy, and so on.

Lastly, what Fitzpatrick suggests is that Mr. Potter "could reasonably have been expected to know better and to do a better job of informing himself morally, given his capabilities and culturally available opportunities."[18] I contend this is exactly where Fitzpatrick takes what is possible to be what is reasonable. From an epistemic point of view, Mr. Potter has never acquired any beliefs throughout his whole life to sufficiently motivate him to know better. The events and mental states of his mind have never provided enough reason for him to act or even to think otherwise. Nevertheless, if we were to assume an indeterministic world, it is certainly conceivable and possible that Mr. Potter chooses differently. It is possible for Mr. Potter to acquire "better" moral judgement without enough reasons, but it is certainly unreasonable to expect him to do so. This is also why the anticipation model does not hold Mr. Potter to be responsible for his action, because the basis of the anticipation model is to have beliefs or anticipations that can reasonably motivate the agent. Asking the agent to choose based not on

[18] William J. Fitzpatrick, *Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge*, 606

what they genuinely believe to be right would certainly be possible, but nonetheless unreasonable.

Another kind of normative ignorance is the failure of completing a "2+2" task, meaning that the agent fails to access stored normative value at the time of action. As mentioned above, the skepticism about responsibility takes this failure to be non-culpable as it is not an akratic act. The anticipation model, in contrast, takes a more detailed approach. There are three possible scenarios that lead to the failure of connecting the general moral belief and the ongoing action. One is that the agent is so limited by factors beyond their control (the complexity of the association is too demanding, the agent was under stress, etc.) that they cannot make the association at the time. The second is that the agent is simply unaware of the connection at the time of action for no particular reason. The last one is that the agent actively chooses not to proceed on the moral reasoning so that he prevents himself from seeing the connection. It seems that we have clearer intuitions for case one and two. We might think that in case one, the agent is utterly unable to see the connection and therefore not responsible, while in case three, the agent is potentially aware of the wrongness and actively attempts to blind himself from fully recognizing it, which grounds for their responsibility. The difficulty arises, however, from case two. Rosen would certainly argue that this is not an akratic act, but our intuition seems to be mixed for cases like these.

Imagine Will, a student who goes to William & Mary, kills a person on campus with a gun. Suppose, epistemologically, there are four possible scenarios that can reflect Will's mental state:

1. Will believes that killing is wrong and realizes shooting causes killing, yet at the time of shooting, Will acts against his better judgment and pulls the trigger.

2. Will does not believe that killing is wrong and firmly believes that shooting is justified. Will has a common upbringing and was not subject to any brainwashing education that causes him to have this thought. Consequently, Will shoots his target with no regret or shame.
3. Will believes that killing is wrong and shooting at a person will most certainly kill him, yet at the time of shooting, Will does not make the association and concludes that shooting is wrong for no obvious reason.
4. Will believes that killing is wrong and shooting at a person will most certainly kill him, yet at the time of shooting, Will does not make the association and does not conclude that shooting is wrong since Will was under severe pressure that disables him from doing so.

According to Rosen, Will is blameworthy only in scenario 1 since it presents a case of true akrasia – Will is consciously aware that his action is morally wrong at the time of action but still chooses shooting. For scenario 2 to 4, however, Will does not think that his action is wrong right before the moment he performs the shooting, negating his blameworthiness under Rosen's view. From Fitzpatrick's perspective, Will is blameworthy in scenario 1 through 3. While Fitzpatrick is in agreement with Rosen that a genuine akratic act occurs in scenario 1 grounds responsibility, he also thinks that Will could have made more effort on self-education so that he has a better moral view in scenario 2. It is Will's own choice that causes him to have an incomplete moral view, i.e., shooting and random killing is justified, which allows Will to be held responsible. Similarly, in scenario 3, there are no constraints that prevent Will from making the association and conclude that killing is wrong. Thus, it is within Will's control to realize shooting is wrong, which is a step that can be reasonably expected.

The anticipation model takes a middle ground. It agrees with Rosen that an akratic act is sufficient for determining blameworthiness. However, in addition to scenario 1, the anticipation model only holds the Will from scenario 3 to be morally responsible.

In scenario 3, if we were to interrupt Will and ask him whether killing is wrong or whether shooting leads to killing, there lacks compelling reasons for us to think Will would answer anything other than what he believes (assume that the interruption does not cause any mental burden). In other words, Will should not have any trouble accessing either belief when asked, given the fact that he has well-internalized both. Had Will had trouble answering either question, it appears that Will is either intervened by other limitations or Will does not have a clear belief altogether. It is certainly possible that Will is not consciously thinking about neither belief at the moment of action, yet Will's mind, contrary to Mr. Potter or in scenario 2, contains necessary and sufficient beliefs for Will to conclude that "shooting is wrong," whereas Mr. Potter could not have made a similar association nor a conclusion at the time of action.

This provokes a similar question presented in the normative ignorance case of Mr. Potter, that is, whether the association can be reasonably expected from Will and subsequently motivates him to do otherwise. It might appear that Will, similar to Mr. Potter, also lacks any reason that motivates Will to make such an association, therefore, one might oppose that Will should be held responsible. Here I present two arguments responding to this comment.

First, I contend if one rejects the notion that simple connection between existing beliefs or ideas can be reasonably expected from Will, one also has to accept that Will has little, if any at all, autonomy over his mind. Because the connection is a simple one, even we assume that there is no additional reason that motivates Will to associate any ideas, it is certainly still within Will's control to do so. Opponents of this view are forced to acknowledge that Will, a free agent,

lacks the ability to modify himself by deliberating between ideas that have already existed in his mind, hence contradicting the premise that Will is a free agent. Unlike Mr. Potter, whose attempt of additional moral self-education is loosely related to changing his moral beliefs, Will's attempt has a much greater, if not guaranteed, chance of leading a new moral view.

Second, from a practical standpoint, it is often possible that Will's failure of making the connection is in contradiction to his other general beliefs, allowing a potential motivator for Will to avoid such a failure. It is generally accepted that one should think about the consequences, both moral and non-moral, of one's action prior to committing to it. A similar idea is that being an adult is to be a responsible individual. These beliefs can act as reasons or even potential causes, other than Will's agency, for Will to make the association. Given the fact that there is no relevant limitations that constrains Will when he pulls the trigger without putting in much thought, we can reasonably argue that Will is culpable for his ignorance.

Accepting that general moral beliefs can be taken as the basis of responsibility solves some issues of using guilt and shame as the evidence of moral wrongdoings. The core challenge mentioned above suggests that, given people's limitation on recalling past, an akratic act is indistinguishable from a normatively ignorant act as people might feel guilt or shame after the act is completed. The anticipation model, by holding certain kinds of normative ignorance to be culpable, does not precisely distinguish whether the act that leads to guilt or shame is akratic or not, but contends that these feelings are linked to one kind of moral wrongdoing or the other with certain confidence.

VI. Speech Responsibility Skepticism

Speech comes in many forms, any kind of expression such as art, movies, or symbols can all be categorized as speech. It has become well-accepted that speech can lead to harmful and

damaging consequences. Former president Trump, for instance, has long mocked wearing masks during pandemic, which some argue fuels the spread of Covid-19 and costs thousands of lives. However, whether the speaker is morally responsible for his speech and the consequences following remains up to debate. Proponents of censorship often treat it as a preventative measure, claiming that it bars much of the damage from occurring. Despite these measures being potentially justifiable, I contend that the results of a speech are irrelevant to one's responsibility. Determining responsibility, as mentioned in the previous chapter, requires access to the speaker's prior anticipation of their remarks rather than knowing the actual consequence. I therefore sketch the following argument for why one should maintain a skeptical position about moral responsibility related to the speech.

As noted, one's normative judgement and general moral beliefs prior to their action has a defining role of their responsibility, which nicely explains our intuition for not holding children or intellectually disabled adults accountable. In general, similar to the legal allegation, the responsibility allegation relies on external evidence that speaks of a person's intent. Consider:

> **Murder:** Rob is charged with murdering his wife, Beth. The prosecutor claims that the CCTV clearly captured Rob stabbing Beth from behind and hide her in their backyard afterwards. The weapon, a kitchen knife, was found to be just bought by Rob earlier in the morning.

Here we have a case where, at the first glance, Rob appears to be clearly responsible for murdering Beth. Though we have no direct information about Rob's thought at the scene, we have crucial evidence that speaks about Rob's mind. According to the prosecutor, the fact that Rob purchased the weapon just before the killing takes place and hide his wife right after suggests he had murderous intent. If the prosecutor were true, Rob had already envisioned or anticipated Beth's death well before it occurred, which ultimately mounts to his blameworthiness.

In Rob's defense, however, he never wanted Beth to die. Rob indicates that he has married Beth for over 20 years, and they are a loving couple known to all their neighbors. Rob also points out that he has long suffered from severe mental issues that causes him to hallucinate. Based on Rob's response, he never recognized Beth on that day. All he saw was a burning demon with a deformed face standing in the Kitchen. Rob's fear prompted him to buy the knife across the street and subsequently kill the demon who is, in fact, Beth. If we assume that it is reasonable and even encouraged for anyone to cast out an intruding monster in their home and suppose Rob is telling the truth, it will indeed make him not blameworthy.

Without further evidence, we and the Judge seem to hit a roadblock. Naturally, we look for additional information such as Rob's medical record and their neighbor's view on Rob and Beth as a couple. We should be concerned about how the additional evidence might only be circumstantial and has no impact on the truth. On the one hand, Rob can have a perfectly fine record but still be schizophrenic since he never seeks medical assistance. On the other hand, Rob could have a long history of hallucinating but not have been ill when he conducted the crime. What Rob saw, in this particular case, is strictly limited in his own head. Fortunately, it is often incredibly hard for one to fake a mental disease given modern technology. Many behaviors do allow us to make reasonable interpretations of the agent's thoughts. Punching, stabbing, or poisoning, to name a few, generally represents ill intentions of the perpetuator and rare exceptions usually require serious medical conditions.

Nevertheless, speech, unlike most of our behaviors, often speaks little about our anticipation. The relationship between speech and its consequences is not as well-established, thus causing the outcome of speech to be less predictable. It is unsurprising how common miscommunications can happen on a daily basis. Words and other forms of expressions can be

badly expressed, wrongly understood, and interpreted without context, all of which ambiguate the true anticipation before the speech. Interestingly, the speaker might not even be consciously aware of what their words can lead to. As a result, I argue that we should maintain a skeptical position towards the moral responsibility of speech. Let us examine a hypothetical scenario:

> **The Psychiatrist**: Dr. Zheng is a renowned physician who has earned her patients' deep trust. She has an impeccable review from her clients and many mentally depressed people, including Lucy, turn to her for help. After a few sessions with Lucy, Dr. Zheng has established a trustworthy figure and becomes Lucy's top confidant. Zheng latter convinced Lucy to commit suicide by demoralizing Lucy.

Dr. Zheng is apparently nonprofessional, but is she as blameworthy as a murderer? Under the anticipation model, if Dr. Zheng genuinely expects her words would cause the death of Lucy in the same way a bullet would end one's life, then for Dr. Zheng, demoralizing Lucy is equivalent to shooting Lucy, which effectively indicates that she is as responsible as directly killing Lucy. Alternatively, if Dr. Zheng believes that directly killing is worse than convincing someone to commit suicide, she would be less blameworthy since she has a different anticipation.

One problem with the above analysis is that it is hardly possible to know Zheng's actual anticipation partly because the consequence of speech is so loosely connected with speech itself compared to other actions. Had Dr. Zheng chose to stab Lucy, we would have much more confidence judging Zheng as a murderer even without knowing Zheng's conversation with her friends. In comparison, although Dr. Zheng has much influence over Lucy due to their physician-patient relationship, we have a rather hard time equating demoralization to physical damage. This pushes us to make inference about Zheng's anticipation based on her first speech with her friend.

We now come to realize that we presume speech mainly has two functions – one that aims to directly put our thoughts into words (and potentially other forms of expression), and the other that merely serves as a means to an end. When Dr. Zheng was being cruel to Lucy, how

much of Zheng's speech is reflective of her own views? Similarly, how much of Zheng's bragging is genuine and can be taken literally? Just like we rely on Rob's medical record to judge his claim about being delusional, we also rely on non-content-related information about Dr. Zheng to decide how much of her speech speaks about her. We try to answer questions like has Dr. Zheng been manipulative towards other patients before? Or how did she deliver her speech to Lucy? Could it be a joke that went wrong? Taken together with the example of Rob, notice that the level of truth of any kind of speech is never gauged by the literal message, but through cues like attitude, manner, or other actions. Speech in itself, however, speaks little about one's anticipation and therefore I argue could hardly be used as a basis for judging moral responsibility.

Many might find this position counterintuitive. Afterall, we are inclined to think that we still retain the capacity, albeit imperfect, to tell the true meaning from literal words or lies. When my grandma tells me that I am not good at philosophy, I have a high level of confidence believing that she simply wants me to improve. When my friends tell me that I am terrible at philosophy, I also know that he is merely teasing me rather than trying to hurt my feeling. I seem to be able to easily attribute moral responsibility to their speech, which renders me with no reason to be skeptical.

However, if we take a closer look at the interpretations I just made, we will conclude that they are not only dependent on the information in addition to the mere content, but also are always subject to changes. The reason why I infer that my grandma is using harsh words to push me to become better is because she has always been a loving character in my life. I trust her to be a kind and warm person based on my experience with her. Moreover, I know that my friend is only making fun of me because he made his comment in a humorous tone, and we also have

been great friends for years. I have much evidence that are not literal to validate my inference, yet the speech itself can hardly be informative. Suppose my grandma makes her criticism with a disappointed look, and she, from that moment on, ceased to be the loving and kind person I knew of. Instead, she became cold and sarcastic, which, in retrospect, I am more likely to believe that she was hurting me with her criticism. Once again, we turn to other clues to form our moral judgement, which can easily be altered based on additional knowledge of the speaker.

At this point, readers may think that, with enough clues beyond the literal information, we are justified to make a judgment on responsibility regarding one's speech. However, in my opinion, these circumstantial clues can only provide negative knowledge of one's anticipation, leading to a mere crude judgement of moral responsibility. Returning to the example of Dr. Zheng, there are two kinds of beliefs she might have – one equates harmful words to flying bullets, and the other holds that talking someone into death is less blameworthy than directly killing. These two anticipations manifest in identical ways: Dr. Zheng remarks a harmful speech to Lucy, leading to her death. While we might be confident that Dr. Zheng does not mean any good to Lucy, we have no extra information to accurately determine which anticipation was in Dr. Zheng's head prior to the speech. In fact, options for Dr. Zheng's anticipation are unlimited. Dr. Zheng can be very confident that her speech will make Lucy commit suicide, or she could be not confident at all. Zheng's levels of confidence play a direct role in her responsibility, yet in manifestation, there are little clues to tell.

Suppose Bob is presented with two pills. Taking the red pill has a 99% chance of mortality rate and taking the blue one only has 1%. Knowing this, Bob chooses to feed me with the red pill over the blue one. Intuitively, he is more blameworthy than feeding me with the blue one. In his pre-deliberation, he foresees that he has a substantially higher chance of killing me

with the red pill. As a result, by choosing the red pill over the blue pill, Bob can be considered as being more serious in killing me and therefore more blameworthy. Unfortunately, I did not know which pill Bob fed me with. In other words, I have little access to Bob's anticipation. Was Bob trying to murder me? Or was he simply putting me under some minor risk? While I am confident that feeding one with the pills above is not a kindness, I am strained to make any further judgment regarding his responsibility. Thus, I am forced to take a skeptical position regarding the amount of blame Bob deserves and suspend my judgement.

Importantly, regarding the responsibility, what matters is what effect Bob thinks the pills have, not what the actual effect the pills have. If Bob misunderstands that the blue pill has a greater probability of killing me and feeds me with the blue pill accordingly, he would still be as blameworthy. This case can be generalized to the issue of speech as well. It is generally known that saying cruel words can harm others, but its impact is up to one's own speculation. Dr. Zheng can anticipate her words to have a 100% chance of killing Lucy, or she can believe that Lucy was put on a very small risk of dying. In retrospect, both anticipations are possible causes of Lucy's death, yet the responsibility of Dr. Zheng drastically differs depending on the belief she holds.

One can accept both the skeptical position and the anticipation, but still believes that counter measures can be taken against wrong doers. While I cannot accurately ascribe the exact blame Bob deserves, I am well-justified to refuse taking the pill or stop him from feeding me with one. This rationale is illustrated through "the Innocent attacker" example. An attacker, Albert, is trying to shoot Victoria, believing Victoria attacked him moments ago. In reality, Victoria's identical twin, Viki, was the actual culprit and Victoria is in fact innocent. Albert wrongly takes Victoria to be Viki and is trying to attack her instead. We have good reasons to

think that even though Albert and Victoria are innocent, the victim, Victoria, is justified to kill Albert to stop him from shooting her if killing is her only viable option. Even if Victoria clearly knows that Albert must have made a mistake, it will still seem to be harsh to blame Victoria on merely defending herself.

Following the same logic, some say, when I encounter someone spreading hate speech against me or issues related to me (such as my race or ethnicity), I am also well-justified to stop them from causing more harm. It is unclear to me what beliefs the speaker holds, but I arguably have good reasons to stop the speaker from making further speech. Based on the anticipation model, as long as I believe that it is morally acceptable for me to defend myself through censorship, I should not be held responsible for silencing the speaker even by force.

Three challenges arise from translating the innocent attacker argument to the personal silencing example above. First, the innocent attacker case presents a scenario of emergency, which does not appear in the personal silencing case. Thus, when better measures are available, say, simply ignoring the speaker, victims should not take extreme measures like forceful silencing. Second, from a third person's point of view, when both parties are innocent, it is unclear which side should be supported. Since many times stopping others from killing or commenting requires a third party such as the police force, proponents of the innocent attacker argument should further discuss why one should prioritize the victim over the attacker. Third, even if measures silencing or censorship were to be taken, to what extent should they be implemented? More specifically, are speakers of harmful speech denied with public platforms only or should they be isolated from those they harmed permanently.

Imagine Veronica is taking the subway to go to work. In the crowded train, she hears a man attacking her with profane words non-stop. With Veronica's noise-cancelling headsets,

Veronica can easily block the harmful speech and go on with her normal routine, rather, she chooses to stop the man by threatening to call the police. Veronica has a clear second option: put on her headsets and not listen to the man, but our intuitive response is that Veronica has a right to stop the speaker (some might even argue that she should stop the speaker, or the speaker should be stopped). From Veronica's perspective, she is not blameworthy for choosing either of the option so long she believes the choice is justified.

The key question is whether hearers like Veronica should have a preference when weighing between options. One might prefer not listening to the speech over stopping the speaker since stopping the speaker causes slight harm to an innocent person and simply not listening does not cause any harm. Though Veronica does not have a clear understanding of the speaker's intention and moral beliefs, it is possible that the speaker is innocent and should not be punished unjustly. Even if we were to assume that the speaker is in fact blameworthy, his blameworthiness can be so slim that it can be effectively considered as negligible. This is because, as reasoned previously, the speaker might not expect his action to have much of an impact and believes that the words he said only causes minor harm. From the speaker's perspective, even though it is wrong to harm others, his speech does not cause much harm and he is only blameworthy to the minimal extent.

Suppose Veronica chooses to notify the police, which leads to the second challenge – what should a bystander, namely the government or the social media, do to resolve the conflict between the speaker and the listener? A more timely example is the multiple permanent suspensions of the former president Donald Trump's social media accounts. Assume that Trump's speech on public platform is a genuine representation of his beliefs, based on the anticipation model, Trump is not morally blameworthy for spreading rumors and hate speech

against minority groups such as Muslim immigrants. However, being a president with wide publicity, Trump's words undoubtedly lead to substantial damage to both the general public and those who he spoke against.

Intuitively, Trump deserves suspensions due to the amount of harm he caused. While both the speaker and listeners are innocent, Trump's beliefs and speech are baseless. From a Thomsonian perspective, he imposes unjust harm against the listeners, which justifies silencing him rather than requesting the listener to actively avoid the speech (Thomson 1991). [19] Trump initially claimed that the cases of Covid-19 will drop "like a miracle" back in March 2020[20]. As a result, combined with his political policies and other media exposure, Trump sent out a message that Covid-19 is not an issue to be concerned with. He also mocked wearing masks, slowing down the progress of containing the pandemic. Since we assume that Trump's words are reflective of his personal view, we can infer that Trump lacks scientific understanding of the pandemic, which prompts him to make inaccurate statements about the coronavirus.

Yet the Thomsonian's view introduces a potential problem of luck by relying on the factual basis of the speech. More often than not, neither the speaker nor the listener has a complete grasp of the objective truth. More importantly, discoveries and truth are often attained through debates. Ultimately, one can only compare which side has better argument or better evidence and subsequently judge which side is more likely to be true rather than actually being true. This poses a threat to the no luck principle. While scientists have a way better chance than Trump of being true, they are, nonetheless, subjective to falsity. One can argue that it is unjust to

[19] Judith Jarvis Thomson, "Self Defense,"
[20] Daniel Wolfe and Daniel Dale, *All of the Times President Trump Said Covid-19 Will Disappear,* CNN (Cable News Network)

punish or prioritize a party based on what happens to be true, which is beyond either party's control.

Perhaps one way of solving this is not to ask what is in fact the truth, but which side presents a more reasonable judgment based on the best possible information. Intuitively, people appear to hold a historical view about one's attitude and speech when judging their moral responsibility. Many cares not only about the content of speech, but how the speaker forms such a view. Racists, for example, are viewed as being more blameworthy in the modern era compared to a century or even 80 years ago. A common explanation to these differential reactions is that it was harder for people to realize the harm of racism as much false justification of racism was taken to be true. People of color were portrayed as being more violent, less intelligent, and satisfactory with their lifestyles as slaves, mitigating slave owners' culpability. As time goes by, all these excuses for believing in racism and slavery were abolished and replaced. One might argue that, assuming a slave owner were to attack a black person, while we may not hold either party to be accountable for their speech and opinion, since the black person has a more cohesive argument with better evidence, we are justified to prefer the listener over the speaker. In a more general term, while we believe that neither party is not blameworthy, one side is clearly more reasonable in their belief. Consequently, we are allowed to prefer one party over another.

On the surface this argument is appealing, yet it raises more questions than it answers. Firstly, it still remains open to debate whether spending more effort or being more reasonable is morally relevant. Secondly, not only is much of the argument based on vague definitions, such as "reasonable" or "best possible." More importantly, a better argument is not definitively associated with the truth. From an individual's perspective, one might know the truth (slavery is

not the best societal system) without understanding the reasons behind it. On the other hand, the opponent might have a much more developed argument with a false conclusion. Imagine a Nazi officer is prosecuting a demented Jewish elderly. The Nazi officer believes that his killing is for the greater good (an idea based on a series justifications) while the Jewish elderly barely retains any rationality. In this case, we are still biased towards the elderly but not the officer.

VII. Defending Freedom of Speech

Historically, the debate about the scope of freedom of speech is often viewed as an ethical question, and some are prone to analyze the costs and benefits of the speech. In his well-known "On Liberty," John Mill defends freedom of speech based on its truth value in an exhaustive manner, arguing that no matter a speech is false, true, or just partially true, it always leads to greater good.[21] Other liberal theorists appeal to concepts of individual autonomy or personal rights; and argue that one cannot be free if one's speech is restricted. Thomas Nagel, for instances, claims an unlimited freedom of speech: "Willingness to permit the expression of bigotry and stupidity, and to denounce or ignore it without censoring it, is the only appropriate expression of the enlightened conviction that the proper ground of belief is reason and evidence rather than dogmatic acceptance."[22] Similarly, Ronald Dworkin, a firm supporter for anti-discrimination, advocates for freedom of expression from a legal perspective. A striking common feature among these philosophers is that they are all well aware of the harm of certain speech, but believes that, all things considered, the risk and harm of free speech is a small, yet necessary, price to pay for a greater benefit. Dworkin's opinion makes a perfect example here:

> "Fair democracy requires…that each citizen have not just a vote but a voice: a majority decision is not fair unless everyone has had a fair opportunity to express his or her attitudes or opinions or fears or tastes or presuppositions or prejudices or ideals, not just in the hope of influencing others (though that hope is crucially important), but also just to

[21] Mill, John Stuart, *On liberty*.
[22] Thomas Nagel, *Personal Rights and Public Space*, 98

> confirm his or her standing as a responsible agent in, rather than a passive victim of, collective action."[23]

Dworkin then suggests that "The majority has no right to impose its will on someone who is forbidden to raise a voice in protest or argument or objection before the decision is taken," effectively arguing in favor of near absolute freedom of speech – "The temptation may be near overwhelming to make exceptions to that principle – to declare that people have no right to pour the filth of pornography or race-hatred into the culture in which we all must live. But we cannot do that without forfeiting our moral title to force such people to bow to the collective judgements that do make their way into the statute books."[24]

What Dworkin, along with Mill and Nagel, argues for is far-reaching. It applies to virtually any kind of expression, even including ones that we intuitively take to be repulsive, such as fighting words, obscenity, and libel. Naturally, the question of whether the scope of free speech can be restricted is provoked, yet the focus of ongoing debates largely centers around the empirical benefits – like protecting minorities or democracy – of speech restriction and neglects its normative property. This has resulted in the popularity of "content-based" censorship, in which the government or public platforms separate the speech from its speaker by only judging the appropriateness of the content. This strategy, in my opinion, oversimplifies the complexities of the normative aspects of speech and excludes many important factors, such as ill-intention, of the speaker. It is crucial to acknowledge that what censorship does goes beyond mere protecting the hearer, it also imposes restriction and thus harm to the speaker. The question has long been left out of the equation is whether the speaker deserves to be censored or punished for their words. Upon my own reflection, I often find that my reaction to the hate speech, along with

[23] Dworkin, Foreword, viii
[24] Ibid.

virtually all kind of speech, is contingent on our perception of the context and the speaker. What we care about extends beyond the content of hate speech itself, instead, what matters the most is whether the speech reflects an intention of harm or culpability.

Our daily interpretation of speech fleshes out a substantial challenge to the content-based censorship. Such a practice focuses the actual harm of the speech and often prioritizes hearers. For example, if a speech raises strong opposition from the public or is a likely cause for a harmful event, say, a riot, it is justified for the public platform to delete the content. There is some merit to this rationale. It certainly serves the purpose of protecting the hearer, who are also often vulnerable, from being further hurt. Yet content-based censorship is inevitably troubled by the problem of consequential luck, a concept nicely put forward by Thomas Nagel. Nagel points out that it is not just for one to be held responsible for factors beyond their control, which includes how things unfold or the consequence of the agent's action. This idea directly charges the legitimacy of content-based censorship, as the meaning of a speech is open to hearer's interpretation. Peter Singer, for example, was once accused of being offensive through his work such as "Speciesism and Moral Status." [25] [26]We have good reason to believe that Singer had no intention of challenging equal right movement, though his work did result in harm. The amount of harm caused by Singer's work highly depends on chancy events like the size of his audience (which can be influenced by other main-stream media), or how his work was presented. More importantly, there is a great chance that Singer is innocent under the anticipation model, as he might firmly believe in his work and judge it to be morally right. Also, unlike the classical case of the "innocent attacker," the conflict between an innocent speaker and a hearer is hardly an emergent one. It seems that a just third party should seek the most impartial way while

[25] Peter Singer, *Speciesism and Moral Status*
[26] Truman Chen, *The Offensive Peter Singer*

minimizing the harm. Through the lens of moral responsibility, it seems like speakers like Singer deserve absolutely no punishment including any form of censorship. Protection of the vulnerable should not be pursued at the expense of imposing undeserved constraints. Even if we were to blame the speaker, their expected harm of his work is still incredibly difficult to be determined. This obstacle further complicates the speaker's exact amount of blameworthiness, challenging the legitimacy of imposing any kind of censorship.

VIII. Conclusion

This thesis intends to provide a normative guidance for moral accountability and apply it specifically to speech. By analyzing the moral responsibility in general and the uniqueness of speech, I offer an intuitively appealing responsibility framework, the anticipation model of moral responsibility, and defend freedom of speech. The anticipation model captures much of our daily responsibility judgement and avoids the issue of resultant moral luck by strictly limiting the basis of responsibility within the agent's control. Specifically, it contends that moral blameworthiness can only be resulted from free acts that either violate the agent's best moral reasoning at the time of action or from free acts that violate the agent's general moral beliefs, under the condition where no positive judgment is available at the time of action. The actual consequence of a committed act, however, is irrelevant to its responsibility. As a result, we should adopt a skeptical position regarding the responsibility of speech under the anticipation model, because speech entails little about the agent's normative judgment, which is the very basis of responsibility. Even the blameworthiness of a speech can be determined, the amount of deserved blameworthiness remains unclear. Therefore, I am in favor of freedom of speech and stand against any form of censorship.

Bibliography

"All of the Times President Trump Said Covid-19 Will Disappear." CNN. Cable News Network. Accessed April 23, 2021. https://www.cnn.com/interactive/2020/10/politics/covid-disappearing-trump-comment-tracker/.

Chen, Truman. "The Offensive Peter Singer." Philosophy Talk, 2017. https://www.philosophytalk.org/blog/offensive-peter-singer.

Dowrkin, Ronald, in Hare and James Weinstein, "Extreme Speech, Public Order, and Democracy," in Ivan Hare and James Weinstein, eds, *Extreme Speech and Democracy* (Oxford University Press, 2009), Foreword.

Fischer, John Martin. "Semicompatibilism." The Philosophers' Magazine, no. 50 (2010): 40–41. https://doi.org/10.5840/tpm20105047.

FitzPatrick, William J. "Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge." Ethics 118, no. 4 (2008): 589–613. https://doi.org/10.1086/589532.

Mill, John Stuart. "On Liberty." Utilitarianism and on Liberty, n.d., 88–180. https://doi.org/10.1002/9780470776018.ch3.

Nagel, Thomas. "Moral Luck." Mortal Questions, n.d., 24–38. https://doi.org/10.1017/cbo9781107341050.005.

Nagel, Thomas. "Personal Rights and Public Space." Philosophy & Public Affairs 24, no. 2 (1995): 83–107. https://doi.org/10.1111/j.1088-4963.1995.tb00024.x.

Rosen, Gideon. "Skepticism about Moral Responsibility." Philosophical Perspectives 18, no. 1 (2004): 295–313. https://doi.org/10.1111/j.1520-8583.2004.00030.x.

Singer, Peter. “Speciesism and Moral Status.” Cognitive Disability and Its Challenge to Moral Philosophy, 2010, 330–44. https://doi.org/10.1002/9781444322781.ch19.

Thomson, Judith Jarvis. “Self Defense.” Philosophy & Public Affairs, 1991, 283–310.

Vihvelin, Kadri. “John Martin Fischer, The Metaphysics of Free Will (Oxford: Blackwell, 1994).” Nous 32, no. 3 (1998): 406–20. https://doi.org/10.1111/0029-4624.00107.

Wolf, Susan. “Sanity and the Metaphysics of Responsibility.” Responsibility, Character, and the Emotions, 1988, 46–62. https://doi.org/10.1017/cbo9780511625411.003.