

The Ethics of Persuasion

Daniel Galley

A dissertation

submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2026

Reading Committee

Colin Marshall, Chair

José Jorge Mendoza

Michael Blake

Program Authorized to Offer Degree:

Philosophy

© Copyright
Daniel Galley

University of Washington

Abstract

The Ethics of Persuasion

Daniel Galley

Chair of the Supervisory Committee:

Colin Marshall

Department of Philosophy

In this dissertation, I use ‘persuasion’ to refer to the provision of reasons offered with the intention of changing the target’s mind on some moral matter. A key premise that underlies much of this dissertation is that engaging others in persuasion constitutes an essential way of recognizing their rational agency. This dissertation consists of three papers, all of which focus on the ethics of persuasion. In the first paper, I argue that we have a *prima facie* moral obligation to *not* forgo persuading others out of distrust in their rational capacity to sincerely engage with moral reasons, since doing so treats them paternalistically. The second paper responds to the claim that attempts at persuasion can risk precluding the target from exercising their rational autonomy by getting the target to defer to the persuader’s judgment. I argue that while attempts at persuasion can result in the target deferring to the persuader’s judgment, the attempt at persuasion does not risk precluding the target from exercising rational autonomy, but rather the attempt risks expressing an ‘authoritative air.’ In the third paper, I make a contribution to a recent argument made by Cecile Fabre (2023), who claims victims of wrongdoings have a *pro tanto* moral duty to accept their wrongdoer’s apology. I begin by identifying a category of wrongdoing that I refer to as ‘complex wrongdoings’, that fall outside of Fabre’s duty to accept apologies. I argue that while the victim is under no duty to accept the wrongdoer’s apology in these cases, the victim can still have a moral duty to sincerely engage with the wrongdoer’s apology as a form of persuasion.

Introduction

The following dissertation consists of three papers. The underlying theme of these papers is the ethics of persuasion. In its basic formulation, I understand the act of persuasion to be the *deliberate attempt* to change the mind of another person on some matter *solely through the provision of reasons*. The act of persuasion is not to be confused with influencing others indirectly. For example, let us imagine that I happen to mention that I recently learned why the sun rises in the east. Let us also imagine that you did not know this and instead thought the sun rose in the west, and that in light of me voicing my realization, you changed your mind on the matter. Have I persuaded you? Although my realization served as a reason for you to change your mind, it seems a mistake to think that it was an act of persuasion, rather than simply an example of inadvertently influencing you with something I happened to share. To avoid an over-inclusive definition that captures all forms of human conversation, I will limit the act of persuasion to require a deliberate attempt to change the target's mind.

We might worry that restricting the act of persuasion to solely the provision of reasons is not inclusive enough, and fails to include non-verbal acts that influence the target that are intuitively thought to fall within the confines of persuasion. For example, let us imagine that I show you pictures of animals that are visibly suffering in factory farms in order to get you to become a vegetarian. Is this an act of persuasion? Despite not giving you verbal reasons, it seems that it is. While pictures and other visual aids might not require verbalization, I think they can still be used to offer reasons when they are offered within the proper context (Rini, p. 17-18). The proper context is simply one where there exists an understanding between persuader and target that non-verbalized evidence is being presented with the intent to change the mind of the target. I will say more about context in the third paper of this dissertation. For now, however, all we need to know is that my definition of persuasion is not limited to the provision of verbalized reasons, but also includes the provision of non-verbal reasons that are communicated to the target within the context of a persuasive attempt.

What we choose to persuade others of can involve a great many things. However, this dissertation will focus solely on persuading people of *moral beliefs*.¹ As we will see, this will encompass political beliefs that are informed by underlying moral values, such as beliefs relating to climate change, vaccine mandates, and immigration policy. While informed by moral values, many of our political beliefs also draw support from empirical evidence. For example, if I want to persuade you to adopt the belief that you have a moral responsibility to limit your greenhouse gas emissions, I might present you with empirical evidence of the exorbitant carbon footprint of people living in the United States. Therefore, moral persuasion can involve persuading the target of empirical evidence that, when presented in isolation, might be viewed as non-moral. We might worry, then, about distinguishing between moral persuasion from non-moral persuasion in cases where persuasive efforts are focused on getting the target to accept seemingly non-moral empirical evidence. However, I do not think these cases should worry us. This is because when empirical evidence is employed in an attempt to persuade the target to change their mind on a moral matter, I take this evidence to serve as providing moral reason(s) to the target in support of a change in their moral belief. Much like how pictures can communicate moral reasons when presented in the proper context, so too can empirical evidence when presented against the backdrop of supporting a moral belief. From here on, when an agent attempts to persuade a target to accept empirical evidence in support of a moral belief, I will refer to such attempts as attempts at moral persuasion.

Throughout this dissertation, I refer to persuasion as “rational persuasion” or “moral persuasion.” I use these terms interchangeably to refer to the provision of moral reasons aimed at changing the mind of the target on some moral matter. A key difference between persuasion and other

¹ Difficult cases to categorize as ‘moral persuasion’ are ones where beliefs are influenced through habituation and/or custom. For example, take Blaise Pascal’s notion that to attain faith in God, one should try to live as if they believed, and then they may come to do so (*Pensées*, p. 74). If I successfully persuade you to try living as if you believed and, after sometime, you eventually begin to believe through habituation and/or custom, have I successfully *persuaded* you to believe and not merely act? In other words, is this a case of moral persuasion? My intuition in these cases is that I would be partially responsible for getting you to believe—if in fact you started believing—but in a somewhat indirect kind of way. With that said, if you did start to believe, I think you would be responsible for much of the doxastic heavy-lifting. It is not clear to me whether successfully persuading someone to act in such a way that brings about the adoption of a belief should be credited to the persuader or the target. For the purposes of this dissertation, however, we need not establish whether adopting beliefs through habituation and/or custom falls under the category of moral persuasion, since we will not deal with such cases.

means of influence, such as manipulation, trickery, and coercion, is that attempts at persuasion utilize moral reasons that the persuader expects the target to rationally engage with. A target rationally engages with moral reasons when they deliberate over reasons with the appreciation of how those reasons support the appropriate moral belief. This expectation of the target to rationally engage with moral reasons is explained by how moral reasons apply universally and categorically, applying to all persons and weighted above individual preferences (Rini, p. 4). While other means of influence like coercion can involve getting the target to rationally appreciate self-interested reasons, attempts at persuasion ask the target to rationally engage with moral reasons as a “user of reasons”, rather than as a mere subject with preferences to be used and controlled (Rini, p. 5). By treating the target as a user of reasons, the persuader expects the target to participate in a practice unique to rational agents of weighing and deliberating over reasons that support a change of mind, and not simply a change in behavior. Other means of influence such as manipulation and trickery do not require the target to rationally engage with reasons, and often benefit from undermining the target’s ability to do so. When these other means of influence are employed, all that matters is that the intended result is achieved. Of course, while not requiring this kind of rational engagement from the target, the other means of influence I have mentioned can still have the power to change the target’s mind. After all, people are manipulated and tricked into acquiring new moral beliefs all of the time. The change of mind associated with a successful attempt at persuasion is unique, however, in that its method requires the target to change their mind *in virtue of rationally engaging with the moral reasons presented*, and not for extraneous reasons, like wanting to please the persuader. This implies that persuasion requires the target to freely engage with moral reasons and maintain the authority to change their own minds based on the reasons presented. Therefore, a primary difference between persuasion and other means of influence is that attempts at persuasion respect the rational autonomy of the target by requiring them to choose to change their minds through free and deliberate rational thinking.

A key premise in this dissertation is that attempting to persuade others can be a way of conveying respect for them as rational agents, capable of engaging with moral reasons. The first paper looks at this relationship between persuasion and respect for rational agency, and specifically at cases

where people decide to forgo persuading the target out of distrust in the target's rational ability to engage with moral reasons and adopt a change of mind. In the first paper, I argue that we have a *prima facie* moral obligation to *not* forgo persuading others out of distrust in their rational capacity to sincerely engage with moral reasons and change their mind. I claim that forgoing an attempt to persuade the target treats them paternalistically when (i) A distrusts B's capacity to sufficiently engage with A's arguments or reasons (ii) and thus avoids trying to persuade B (iii) because A assumes his judgment of B's competency to engage in moral matters is superior. Part of what appears insulting with interpreting the target as rationally incapable of engaging with our moral reasons is that it involves taking an "objective attitude" toward them by treating them as a thing to be managed, in that it writes off what they believe as something not to be engaged with on its own terms. In cases of moral and political disagreement, I argue that we ought to attribute the target's belief to a product of rational *error* rather than rational *incapacity*.

The second paper responds to the idea that persuasion can sometimes serve as a tactic that gets the target to act and think in a way the persuader chooses, rather than letting the target decide for themselves (Price, 2020). On this view, persuasion runs the risk of precluding the target from exercising their rational autonomy by getting the target to defer to the persuader's judgment, rather than letting the target decide for themselves. I argue that while attempting persuasion can sometimes result in the target deferring to the persuader's judgment, it can be morally problematic not because it risks precluding the target from exercising rational autonomy, but rather because it risks expressing an 'authoritative air.' I claim that the mistake in thinking that persuasion precludes rational autonomy comes from associating it with what I refer to as 'blunt deference', defined by cases where A risks getting B to defer when A tells B, "you should believe X because of Y," without an expectation that B rationally appreciates why Y supports X. I claim that, by definition, persuasion cannot elicit blunt deference, since the target's rational appreciation of supporting reasons is necessary to its aim. Alternatively, persuasion can elicit what I refer to as 'subtle deference', defined by cases where A tells B, "you should believe X because of Y," with an expectation that B rationally appreciates why Y supports X, but without leaving any room for B to argue back (A might say, "look, stop arguing and just think

about what I've said - you'll understand why I'm right"). The upshot of my analysis is that it explains the concern with persuasion in terms of a failure to reciprocate respect, avoiding the counterintuitive idea that persuasion precludes rational autonomy.

The third paper contributes to the literature on both persuasion and apologies by adding to a recent argument made by Cecile Fabre (2023), who claims that victims of wrongdoings have a *pro tanto* moral duty to take wrongdoers as being sincere and accept their apology. In this paper, I identify a category of wrongdoings that I refer to as 'complex wrongdoings', that fall outside of the duty to accept apologies. I claim that in these cases, the victim has relevant and reliable evidence to doubt the sincerity of the wrongdoer's apology. I claim that in cases of complex wrongdoings, apologies can constitute an attempt at persuasion. I then explain the Kantian duty that targets of persuasion have to open themselves up and engage with an agent's persuasive attempt (Rini, 2018). In cases of complex wrongdoings, I argue that while the victim is under no duty to accept the wrongdoer's apology, the victim can still have a moral duty to sincerely engage with the wrongdoer's apology as a form of persuasion. The upshot of my argument is that while with many wrongdoings victims will not have a duty to accept their wrongdoings apology, they might still have a duty to engage with it. This paper contributes to the literature on apologies by demonstrating the key role persuasion has in the ethics of apologies.

The broader contribution of this dissertation can be understood amidst the fairly recent criticisms of the use of persuasion in moral and political matters. Some of these criticisms begin by pointing out the difficulty in resolving deep moral and political disagreements, evidenced by the ongoing failure to find an equilibrium on matters such as climate change, immigration policy, and vaccine mandates. Helping to perpetuate this failure is the often futile effort from people to attempt persuading their moral and political opposition to accept their view. Critics claim that this effort is often futile given the fact that deep moral and political disagreements are seldom resolved through persuading others with facts, reasons, and arguments (Haidt, 2001; Kahan and Braman, 2006). But not only do critics think that persuasion is generally ineffective at resolving such disagreements, they also claim that efforts to persuade others can be unethical, as they can exacerbate existing conflict by coming

off as righteous, invasive, condescending, paternalistic, and forceful (Haidt, 2013; Tsai, 2018; Price, 2020). This dissertation contributes to the literature on persuasion by resisting the general notion that we should abandon attempts at persuasion in the face of deep moral disagreement. Instead, I call on us to re-examine our notion of ‘successful’ persuasion in terms of the respect it conveys for people as rational agents.

References

- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.
<https://doi.org/10.1037/0033-295X.108.4.814>
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Pantheon/Random House.
- Kahan, D. and Braman, D. (2006) Cultural Cognition and Public Policy. *Yale Law & Policy Review*, 24(87) 147-171.
<https://ssrn.com/abstract=746508>
- Pascal, B. (2018). *Pensées* (W. F. Trotter, Trans.). Dover Publications, Inc.
- Price, T. L. (2020). *Leadership and the ethics of influence*. Routledge.
<https://doi.org/10.4324/9780429449161>
- Rini, R. (2018). Abortion, Ultrasound, and Moral Persuasion. *Philosophers' Imprint*, 18(6): 1–20.
<http://hdl.handle.net/2027/spo.3521354.0018.006>
- Tsai, G. (2014). Rational Persuasion as Paternalism. *Philosophy & Public Affairs*, 42(1): 78-112.
<https://doi.org/10.1111/papa.12026>

You Can't Handle the Truth: Paternalism and Moral Persuasion

Introduction

David is a university professor traveling to give a lecture on the challenge of mitigating the spread of misinformation surrounding vaccines. He decides to stop at a diner for coffee and ends up striking a conversation with a waitress named Rebecca. David learns that Rebecca is five months pregnant and plans to decline all vaccines for her child. She strongly believes that the science around vaccines is false, and that the falsification of the scientific information is part of the governments' effort to exert control over, and exploit its citizens. David is concerned for Rebecca's unborn child, and is eager to show Rebecca the evidence that supports the safety of vaccines, as well as the reasons for why a government coverup is highly unlikely. After all, this is exactly the argument he will give during his lecture. He stops himself, however, because he realizes that Rebecca's position on the matter is irrational, and takes her capacity to weigh any new reasons and evidence that bear on the matter to be deficient. Unlike the students and faculty who will be attending his lecture, trying to persuade Rebecca would be a lost cause, David thinks, and would only lead to misunderstanding and frustration.

Deep disagreement over moral and political matters like the one illustrated above seem to rarely be resolved by one side trying to rationally persuade the other of their mistaken viewpoint. I use the term 'rational persuasion' to refer to the provision of reasons communicated to the other with the intention of changing that person's mind on some moral or political view they hold.² Disagreements like these seem to rarely be resolved by dumping facts, arguments, and reasons onto your interlocutor. But what about alternative approaches? One alternative is to avoid engaging your interlocutor on the matter altogether, as we see David does with Rebecca. Another is to avoid initial disagreement in order to earn the trust of your interlocutor before you decide to direct arguments at them. Although

² Of course, rational persuasion need not be limited to moral/political matters. However, for the purposes of this paper, I use the term to refer exclusively to such matters.

appealing alternatives for avoiding unproductive conflict, my concern with these alternative approaches lies in the strategic attitude we are prone to take toward our interlocutor when employing them. My focus, then, is not on which methods of persuasion prove most effective at changing the mind of our interlocutor, but rather on what we owe our interlocutor in terms of respect. The case of David and Rebecca, I claim, is characteristic of how forgoing rational persuasion can be paternalistic. I argue that we have a *prima facie* obligation against forgoing rational persuasion out of distrust for others' capacity to form the correct judgment on moral matters.

This paper will proceed as follows: In section 1, I build on Seanna Shiffrin's (2000) motivational account of paternalism and argue that agent A treats agent B paternalistically when (i) A distrusts B's capacity to sufficiently engage with A's arguments or reasons (ii) and thus avoids trying to persuade B (iii) because A assumes his judgment of B's competency to engage in moral matters is superior. In section 2, I give an account of how the persuader should interpret the target charitably by interpreting the target's problematic position as a product of rational *error* rather than rational *incapacity*. In section 3, I look at George Tsai's position on how rational persuasion can treat the target paternalistically and explain how it presents a challenge for my view. I respond to this challenge by explaining how Tsai's examples of paternalism rely on factors extraneous to the method of rational persuasion. Lastly in section 4, I explain two types of cases that serve as exceptions to the obligation to engage one's interlocutor in rational persuasion.

I take my paper's contribution to be twofold: (1) While the bulk of the paternalism literature focuses on the conditions and motivation behind *actions* (Begon, 2016), my paper contributes to the literature by looking at the motive of paternalism behind *inaction*. As I will explain, there are certain types of respect that we owe to others and so when we forgo extending these it can meet the conditions of paternalism. In these cases, it is not the intervention that constitutes the insult, but rather the motive of distrust the persuader has for the target. (2) The other contribution of my paper comes from shedding light on the importance of rational persuasion not as a tool for influence, but more importantly as a valuable method for conveying respect. Some of the literature in philosophy and moral psychology (Haidt, 2012, Tsai, 2014, Price, 2020) raises moral and practical problems with the

attempt to rationally persuade others of political and moral views. By comparison, my paper serves to defend rational persuasion as a desirable method of influence. By attempting to rationally persuade another agent, the persuader respects the target as a moral agent by trusting in the target's capacity to rationally deliberate. While there have been issues raised by other authors, my paper looks at the respect conveyed to others when deciding to engage them in rational persuasion.

1. When forgoing moral persuasion is paternalistic

In the previous case, we might find it peculiar to think that David treats Rebecca paternalistically given that he never actually *intervenes*. On the contrary, he stops himself from doing so. Typically, philosophers think of paternalistic treatment as occurring through some sort of intervention made on another (Begon, 2016). Moreover, paternalism is typically thought to be motivated by concern for the other's best interest or welfare. David's motivation to forgo persuading Rebecca, however, is based purely in a distrust for Rebecca's rational capacities, and nothing more. In this section, I will address these peculiarities. I will do so by drawing on Seana Shiffrin's influential account of paternalism, and show how her account helps to explain why the insult with forgoing rational persuasion can be understood specifically as a paternalistic insult. As we will see, on Shiffrin's account, paternalistic treatment can occur through inaction, and also need not be motivated by concern for another's best interest or welfare. Building on Shiffrin's motivational account of paternalism, I will explain three features of the opening case that make David's treatment of Rebecca paternalistic. I will argue that that agent A treats agent B paternalistically when (i) A distrusts B's capacity to sufficiently engage with A's arguments or reasons (ii) and thus avoids trying to persuade B (iii) because A assumes his judgment of B's competency to engage in moral matters is superior.

In an example given by Shiffrin (2000), she claims that person A treats person B paternalistically when A refuses to help B with a carpentry project because A thinks that B too often depends on others for help and believes B would benefit from learning how to work independently. And yet,

If A voices these reasons and persuades B to do it himself, her persuasive efforts and her subsequent abstention would not be paternalist. But were A just to refuse baldly to help for those reasons, then it seems to me that A's omission would be paternalist (p. 213).

Shiffrin's case illustrates how omissions of action can be paternalistic.³ The case also helps to explain how David's omission in his decision to forgo persuasion can be paternalistic.⁴ According to Shiffrin, what makes A's treatment of B paternalistic is that A's mere refusal bypasses B's agency in that A inserts her own judgment in place of B's on a matter concerning his interest. On the other hand, by attempting to persuade B with reasons, A engages his agency by allowing him to exercise his own judgment on what is in his best interest. But we might notice that even if A persuades B, A still seems to distrust B to secure his best interest, hence A's effort to persuade him. The only difference now is B becomes aware of A's distrust, whereas he does not if A keeps it to herself. And so the objection might go as follows: what is paternalistic in both refusal and persuasion is A's distrust of B to secure for himself what A thinks is in his best interest. Therefore, A's persuading B does not make her distrust any less paternalistic, but perhaps even more insulting now, since B is made aware of A's distrust of him.

I think the difference between refusal and persuasion can be explained by considering a distinction between A's distrust of B's *will* and A's distrust of B's *judgment*. As Tsai (2014) puts it, "In distrusting B's agency, A may distrust B's judgment (capacity to judge correctly what is in [his] good), or B's will (capacity to act effectively to practically implement or secure [his] good), or both" (p. 86). Where Tsai uses this distinction to catalog different kinds of paternalism, I think it helps explain the difference between refusal and persuasion in Shiffrin's case. That A distrusts B's *will* to secure his interests is indicated both in refusal and persuasion. In refusal, A aims to change B's behavior by circumventing his will. In the case of persuasion, although A's attempt to persuade B does not

³ See McGrath (2022) for a reading of Shiffrin that supports a view of how charitable interpretations can be paternalistic.

⁴ We might notice that unlike in the case of David and Rebecca, the paternalistic treatment in this case is motivated by A looking out for the interest of B. However, later in the section I will explain that Shiffrin's account of paternalism does not require that paternalistic treatment be motivated by looking out for the interest or wellbeing of another.

circumvent B's will, it does seem to supplement his will, at least initially, by offering reasons aimed to help him implement his best interest. The difference, however, is that when A attempts to genuinely persuade B, A places her trust in B's *judgment* by appealing to him with reasons that A expects B to weigh and consider correctly against his own. In other words, A's trust in B's capacity to judge correctly his best interest is implicit, at least to the extent in which A thinks she can successfully persuade B with reasons.

What makes A's mere refusal different from A's persuasion, then, is that she supplants B's judgment for her own in the former, enacting her distrust of B's ability to correctly weigh and consider her reasons against his own. To help motivate the point of how persuasive efforts can respect the *judgment* of an agent but fail to respect their *will*, consider a family intervention for an alcoholic named Tim. That Tim's family has organized an intervention suggests a perceived failure on the part of Tim to look after his best interest, and thus signals a distrust in his will. However, organizing the intervention also suggests that Tim's family holds out faith in Tim's capacity to judge correctly that his drinking has become a detriment to his well-being. On the other hand, Tim's family seems to distrust both his will and his judgment if they viewed the intervention as a lost cause, and believed Tim to be incapable of realizing his drinking problem.

Returning to the case with which we started, we might notice that, unlike with the case Shiffrin gives us, David never openly *refuses* to persuade Rebecca, as he is never given the chance to do so. However, omissions can be paternalistic even when they are not overtly presented, as in the case of a refusal. This is because with both types of omission, refusals and covert omissions, the other person's judgement can be substituted for the omitting party's. For example, take a version of the previous case where Tim's family decides not to intervene because they think Tim is incapable of correctly judging that his drinking has gotten out of hand. Since Tim never asks for an intervention, Tim's family never actually *refuses* his request. And yet, because they distrust his ability to judge correctly on matters that concern his best interest, Tim's family substitutes Tim's judgment for theirs when they decide not to intervene. Therefore, it can be paternalistic of David to forgo rational persuasion no matter whether Rebecca asks to be persuaded or not. This is because David's decision to not persuade her is based in

distrust of her rational capacity to engage with his reasons, whereby he substitutes her judgment of her rational capacities for his.⁵

The first feature of David's paternalist treatment of Rebecca is that he distrusts her judgment in normative matters concerning vaccines. That is, David distrusts Rebecca as a deliberative interlocutor, capable of correctly engaging with his reasons and evidence that bear on her judgment concerning vaccines. That David plans to provide the very same reasons and evidence to those attending his lecture indicates that his distrust of Rebecca is not meted out equally, but is borne from learning she holds specific beliefs about vaccines. Now we might wonder whether David's distrust is also due to factors outside of Rebecca's beliefs, such as her position as a waitress. As an academic, perhaps David shelters implicit bias toward non-academics as intellectually inferior, and so considers Rebecca's beliefs less favorably than he would have had they been expressed by a student or colleague attending his lecture. Apart from intellectual bias, there are a variety of other biases David might possess that could possibly contribute to his distrusting her judgment, and so the decision to forgo persuading Rebecca could be motivated by factors external to her views about vaccines. With that said, I want to put aside external factors and instead focus exclusively on the distrust an agent can have for another's judgment *because* of the moral/political beliefs they hold.⁶ Although external biases are surely worthy of ethical consideration, I think David distrusting Rebecca's judgment in light of learning about her normative beliefs is significant, and will be the focus in the following.

Normative beliefs influence the way we interact with others and move through the world, playing a significant role in shaping our social identities. Because our normative beliefs significantly impact the way in which we relate to others, it is important to give careful consideration to the attitudes we adopt toward others in light of learning about their moral and political belief systems. Accordingly, holding certain moral/political beliefs can have a notable impact on the way a person is treated and seen by others. For instance, if you believe that the earth is flat, you are not simply

⁵ As I will explain later in this section, the decision to forgo persuading Rebecca is paternalistic even though the decision is not motivated by what is in Rebecca's best interest. For now, I hope to have established that omissions can be paternalistic even when they are not overtly presented.

⁶ Additional external factors include, but by no means are limited to, scenarios in which David forgoes persuading Rebecca because he rather enjoy his coffee than argue, or is pressed for time, or is fatigued and simply does not want to debate ethics, etc.,

considered misinformed as a child is thought to be when believing in Santa Claus, but rather are thought to be deeply delusional and perhaps irredeemably so. And yet, similarly to a child, putting stock into such beliefs will likely cause you to be perceived as an incompetent judge in deliberating about such matters.⁷ Of course, sectioning off others from moral discourse can sometimes be justified, and may be necessary in cases where one cannot reasonably be expected to reliably respond to the reasons given them, such as in the cases of children and persons with certain mental impairments. And yet, such cases are typically thought to be unproblematic for the same reason we think it is unproblematic to not hold these same persons fully morally responsible. That is, when a person cannot be reasonably expected to adequately respond to reasons, we typically think that said persons are less morally responsible than persons capable of rational deliberation. Although such persons demand our respect, their status as moral agents is contained to the role of, what Kennet and Fine (2009) refer to as *reason trackers*, where reasons are registered only at a cursory level, rather than reflectively as reasons qua reasons. Examples of reason tracking can be found in the gut feelings we sometimes get, cueing us into attending reasons without revealing to us their content. While tracking reasons can motivate action and guide behavior, a defining feature of moral agency excluded in mere reason tracking is the ability to deliberate on, and reflectively endorse moral judgments, performed by what the authors refer to as *reason responders*. That we view ourselves and others as reason responders is what obligates us to recognize and receive the moral demands of others, and is often used to explain why children and cognitively impaired persons do not always warrant the same demands. Failing to recognize others as reason responders would thus neglect to account for a significant feature of moral membership.

The second feature of David's paternal treatment of Rebecca is that he supplants her agency when he enacts his distrust of her judgment by deciding to forgo rational persuasion. David views Rebecca incompetent to correctly engage with the facts, evidence, and arguments in deliberation on vaccines, causing him to view any attempt to give her reasons as futile, regarding her as a lost cause. As stated at the outset, I use the term rational persuasion to refer to the provision of reasons communicated to the other with the intention of changing that person's mind on some moral or

⁷ Moreover, because condescension toward a belief or set of beliefs risks targeting not only the view in question but also the person holding the view, we may be excluded from other domains of deliberation as well.

political view they hold. The function of rational persuasion, then, can be understood as not merely aiming to change a person's behavior, but rather changing a person's belief or view (Rini, 2018). For example, consider a scenario where I try persuading you to recycle your plastic because I believe that it is unethical not to do so. Let us say you disagree with my view. In one scenario, I could try convincing you by saying, "you better recycle...or else." My coercive speech here suggests that my primary concern is not to change your view about recycling, but rather to get you to start recycling your plastic, even if it is for the wrong reasons. In another scenario, I attempt to persuade you to recycle by explaining to you why recycling is ethically responsible, and that not doing so is to unnecessarily harm the environment. Here, it matters to me that you engage with my reasons, and not simply recycle because it will, say, make me happy. When I try to rationally persuade you, it matters to me that you sincerely engage and consider my reasons with an openness to the possibility of changing your mind, rather than simply changing your behavior.

Where cases of coercion and manipulation generally can be distinguished from cases of rational persuasion with little difficulty, there are alternative approaches that are less clearly distinguishable from rational persuasion, but are nonetheless importantly different. For instance, although David distrusts Rebecca to arrive at the correct judgment through reason and argument— rather than choosing not to engage at all— he may instead choose to take a more tactful approach by employing methods such as "pandering" and "friendliness" (Ferkany, 2021). Instead of trying to convince your interlocutor of the errors in their view, pandering is an approach that appeals to the value system of your interlocutor by attending to their concerns and what they find most important. For example, if you want to persuade someone to become vegetarian, appeal to them with evidence that shows how eating meat affects something they are already concerned with. On the other hand, the approach of friendliness entails adopting a sympathetic attitude toward the other's position in order to elicit a reciprocated understanding from them. Because deep disagreement can often become hostile and uncaring, communicating friendliness can help soften the combative guard that seems to inhibit open-mindedness and charitable interpretation. Although pandering and friendliness can be favorable in affecting change to a person's views, they presuppose a distrust for an agent's rational judgment by

circumventing perceived defects thought to make persuading the agent less likely. Similarly to why manipulation is often thought objectionable, pandering and friendliness are methods used to exploit a person's deliberative capacities instead of promoting them. What makes these methods difficult to distinguish from rational persuasion is that they typically still involve some expectation from an agent to respond appropriately to the reasons provided, thus implying at least some degree of trust in their capacity to make the correct judgment. Though perhaps then not as distrustful as forgoing persuasion altogether is, pandering and friendliness can be distinguished from rational persuasion by the degree to which the former is motivated by distrust in rational judgment of an agent.

The third feature of David's paternal treatment of Rebecca is that David's distrust is directed at moral matters that belong to Rebecca's sphere of control. The majority of paternalistic accounts have focused on the insult involving unsolicited concern for another agent's welfare, in which distrust for an agent's capacities leads the paternalist to interfere with matters of welfare he or she has no authority over (Begon, 2016). However, Shiffrin has notably argued that the scope of paternalism extends beyond matters of a person's welfare, including matters that rightly belong to a person's sphere of control. A significant reason for why we typically think interfering in someone else's welfare can be paternalistic is that it seems disrespectful to interfere with an area of a person's life that they themselves are expected to have control over. The point is that there are other areas in addition to personal welfare that seem also to be within our autonomous jurisdiction, making intrusion on them disrespectful in a similar way to intruding on welfare. Although determining what is rightfully within a person's sphere of control can be difficult to determine, Shiffrin offers a helpful example of an attendee at a talk who has been called on to speak but is interrupted when,

an excited, sympathetic colleague loses self-control and interjects: "Isn't this a better way to put the point?" She goes on to drown him out while cleverly and eloquently articulating his point. She takes over his question because she feels she has a better command of it than he does" (p. 217).

The paternalist intervention in this case need not be motivated by concern for the attendee's welfare, as we can imagine the colleague is rather concerned with the quality of the talk and interjected because she believes the attendee did not have an adequate grasp on the point he was trying to make. If you have ever had the experience of someone interjecting before having finished your thought, you may be familiar with the patronizing feeling of having someone else assume authority over the communication of your ideas. A significant reason for why it can feel insulting when someone interferes with your welfare is because what constitutes your best interest seems to be chiefly *your* concern. I contend that a person's capacity to arrive at the correct moral judgment is also a component of one's agency that they themselves should be recognized as having authority over. Much in the same way we think of one's own welfare, there is a certain responsibility we hold persons to when forming the correct moral judgment. That fully autonomous agents are recognized as primarily responsible for their own welfare and moral judgments is indicative of why supplanting one's agency in either area can be paternalistic. Therefore, even though the cases I have in mind—characteristic of David and Rebecca's—may not entail the intrusion of someone else's welfare that most accounts of paternalism entail, I take them to be similarly paternalistic in their intrusion of someone else's capacity to rationally engage with moral reasons.

2. What we owe our interlocutor

As stated at the outset, I am concerned with the question of what we owe our interlocutor in terms of respect. So far I have argued that David adopts a paternalist attitude toward Rebecca by interpreting what she says as indicative of her incompetence as a moral judge in the normative matters pertaining to vaccines. The purpose of this section is to clarify the obligation David has against forgoing rational persuasion. That is, I look to answer the question of how David should interpret Rebecca's position on vaccines. As I will explain, the obligation will require David to do more than simply engage in rational persuasion. I argue that he must charitably interpret Rebecca's moral position as a product of rational *error* rather than rational *incapacity*.

To use Strawson's (1962) terminology, part of what appears insulting with David's interpretation of Rebecca's position is that it involves him taking an "objective attitude" toward her by treating her as a *thing* to be managed, in that he writes off what she has said as something not to be engaged with on its own terms, but rather as the product faulty judgment. According to Strawson, interpreting the actions of an agent with an objective attitude entails viewing said actions as products of non-rational causes, whereby we suspend viewing an agent as a suitable recipient of "reactive attitudes", such as blame, anger, and praise. The problem, then, is not that David is frustrated or angry by realizing that any reconciliation with Rebecca's view on vaccines is impossible, but it is rather that he views any attempt to get her to understand and respond to his reasons to be futile. In this sense, David ceases to view Rebecca as an agent and instead views her as a thing, as he adopts an attitude toward her that treats her as something to be managed, manipulated, and not engaged with at all.

The attitude David takes with Rebecca is an objective attitude, and not merely an attitude that attributes partial rationality to her, since David thinks of Rebecca as *entirely* incapable of engaging with his reasons, and not merely *lacking* in reasoning ability. On the other hand, it would be an attitude attributing partial rationality to her if David said to himself, 'I can tell that Rebecca doesn't have the scientific knowledge about vaccines that I possess, but she will come to understand my view given an informal framing of the evidence.' In this case, by trying to appeal to Rebecca's rational capacities despite her perceived rational limitations in the matter, David's attitude would attribute partial rationality to Rebecca and not constitute an objective attitude. As I will explain later in this section, having partial distrust in the target's reasoning ability on a given matter is not objectionable, but rather a necessary precondition for the attempt at persuasion. For now, however, I want to focus on the objectionable cases where a persuader takes an objective attitude toward the target by writing off what she has said and viewing her as rationally incapable of engaging in a matter altogether.

As Mark Schroeder (2019) explains, interpreting a person as a *thing* rather than a rational participant can diminish that person's contributions significantly. To illustrate, Schroeder imagines a scenario in which his wife, Maria, compliments him on the jasmine vine he has been cultivating. Upon receiving his wife's compliment, however, he wonders to himself whether her praise is genuine, and not

instead a side-effect of her cheery mood after finding a coin in the vending machine earlier that day. As Schroeder explains,

It diminishes Maria for me to view her favorable opinion of my jasmine vine, which, after all, is still coming along only slowly, as the product of her environment, rather than as a sincere judgment about horticulture to be engaged with on its own terms—a compliment to be accepted with gratitude or demurred from, disagreed with or enthusiastically concurred with (p. 97).

Viewing Maria's compliment as the product of non-rational causes rather than a genuine statement is to treat her as a thing, subject to the whims of her environment and nothing more. In order for Schroeder to view Maria's compliment as constitutive of what she actually believes, he must view her as an appropriate recipient of the reactive attitudes that mere *things* do not warrant. Where it is nonsensical to become angry with a mere thing, such as an acorn that lands on your head, a defining quality of agency is that it includes a participatory role in which actions and decisions can be intended and not merely caused, reacted to and not simply managed.

However, Schroeder suggests that interpreting persons as things is not always diminishing, but can facilitate engagement with others as rational agents—or to use Strawson's language, when it enables taking the 'participant stance.'⁸ To illustrate, Schroeder considers a case in which Maria disapprovingly remarks that his jasmine vines are growing too slowly. This time, Schroeder wonders to himself whether his wife's comment is the result of a bad day at work, rather than an expression of what she really thinks about the growth of his jasmine vines. Therefore, similarly to the first example, Schroeder entertains a causal interpretation of his wife's comment that is not constitutive of her agency. And yet, unlike in the first example, Schroeder thinks that interpreting Maria's disapproving remark in causal terms is not only *not* diminishing, but is also an interpretation that enables the participant stance. As Schroeder explains,

⁸ Schroeder acknowledges that a similar view can be read in Strawson, who suggests that both occasionally employing the objective attitude is inevitable, and doing so can help facilitate the participant stance.

...deploying causal interpretation as an interpretive tool is actually one of the key elements of successful interpersonal engagement. In order to most successfully engage with another specifically as a person—or put differently, in order to fully occupy the participant stance toward them—you must have the tool of causal interpretation at your disposal (p. 105).

What explains the asymmetry in these cases? When does causally interpreting someone's behavior enable engaging someone within the participant stance, instead of diminishing them? Schroeder's explanation is that people are constituted by the most *charitable* interpretation of their behavior, and because the participant stance requires engaging with the behavior constitutive of that person, causally interpreting someone's behavior enables the participant stance when it is the most charitable interpretation available. Drawing an analogy to the principle of charity in philosophy, Schroeder explains that in the same way it is often thought that a charitable interpretation is the one that best represents an author's text, so too is the case with charitably interpreting people's behavior. In other words, Schroeder suggests that not only is the charitable interpretation of behavior non-diminishing, but is also epistemically unproblematic in that charity tracks the truth of a person's behavior in the same way it tracks the truth in the contribution of a philosophical text.

Returning to the case of David and Rebecca, I want to argue that a similar interpretive norm to that of Schroeder's should be applied by David—as well as by others in resembling cases—to avoid treating Rebecca paternalistically. One difference from Schroeder's interpretive norm is that David's charity will not come from a causal interpretation of Rebecca's judgment. Rather, his charity will come from interpreting her as having rational limitations in moral matters concerning vaccines. As previously explained, the attribution of partial rationality is different from taking the objective attitude in its degree of attribution. While the former attributes partial rationality to an agent in a given area, the latter attributes none. In what follows, I will explain why having a minimal distrust in Rebecca's reasoning about vaccines is unobjectionable. Schroeder's example of charitably interpreting his wife

Maria helps to illustrate how David can interpret Rebecca's judgment on vaccines charitably through a partial distrust of her reasoning.

As stated earlier, a significant reason for why I think David's treatment of Rebecca is paternalistic is because he distrusts her judgment entirely on the matter of vaccines. Distrusting a person's capacity for judgment, as I have used the term, is to distrust a person's capacity to arrive at the correct conclusion in light of the reasons provided. The kind of distrust that I have in mind is local rather than general, meaning that it is confined to distrusting a person's judgment on a particular matter or topic rather than a wholesale distrust of a person's judgment across all matters. Where Schroeder is prevented from the participant stance when he takes Maria's critique as constitutive of her judgment, I think David is similarly prevented from taking the participant stance with Rebecca when he interprets Rebecca's view as illustrative of her incapacity to arrive at the correct judgment on the matter. In Schroeder's case, interpreting Maria's critique of his jasmine vine as a product of a bad day at work causes him to write off what she has said as a non-rational reaction, thereby treating her as a thing to be managed rather than a person to be engaged with. In David's case, interpreting Rebecca's remarks about vaccines as reflective of her incapacity to judge correctly on the matter causes him to forgo rationally persuading her, and thus prevents him from engaging with her as a rational agent.

As Schroeder points out, "rational persuasion is a move within the participant stance", since unlike other means of influence such coercion and manipulation, to rationally persuade you entails recognizing you as capable of changing your mind in response to reason (p. 98). It is precisely this recognition of you as a reason responder that makes you an appropriate candidate for reactive attitudes such as disappointment, gratitude, praise and blame etc.. Accordingly, any genuine attempt at rational persuasion will entail engaging you with the participant stance. But in order for David to engage Rebecca in rational persuasion, he must first sufficiently trust her capacity for judgment. As Tsai states, "One has to relate to the other with the expectation that she is capable of forming the correct judgment in light of reasons one provides her, and indeed rely on her to exercise that capacity (as it is meant to be exercised)" (emphasis mine, p. 90). As Tsai points out, one of the key components of rational persuasion that makes it different from manipulation is that it functions to change the mind of the

subject by promoting rational decision-making, rather than undermining it. With this comes an expectation from the persuader that, if the right sort of evidence and reasons are provided, the target will exercise their capacity to arrive at the correct judgment. Underlining this expectation is a certain degree of trust in one's interlocutor that David fails to have in Rebecca when he sees her as a lost cause. Distrusting a person's capacity to arrive at the point you want them to arrive at is incompatible with the decision to rationally persuade them. In other words, it only makes sense to persuade your interlocutor of some view if you sufficiently trust them to arrive at that view after they are given the right sort of reasons. Therefore, the decision to rationally persuade will imply at least a minimal trust in the persuadee's capacity to arrive at the correct judgment.

How might David charitably interpret what Rebecca has said in a way that enables him to take the participant stance? I think the answer is for David to interpret Rebecca's belief of vaccines as a product of rational *error* rather than rational *incapacity*. In other words, David should attribute Rebecca's view of vaccines to a specific fault in her reasoning process, rather than to an incapacity to judge correctly on the matter of vaccines altogether. The difference is that the former is to view the target as incapable of arriving at the conclusion supported by the reasons you provide them, whereas distrusting a specific point in someone's reasoning is to view them as having made an error—or an error one is likely to make if left to their own devices—in route to a conclusion they otherwise would not have arrived at had they been provided the right sort of reasons.

To see how this sort of distrust is compatible with the participant stance, imagine an alternative scenario where David interprets Rebecca's view of vaccines as the result of her failing to gather reputable evidence, instead of writing her off as a lost cause, incapable of arriving at the correct view. In this case, David's distrust would not prevent him from rationally persuading Rebecca anymore than it would prevent him from persuading his audience at his lectures on the importance of recognizing and gathering credible scientific data on vaccines. In both cases, as long as there is at least a minimal trust in the persuadee's capacity to arrive at the correct judgment, a distrust that the persuadee has reasoned correctly about a matter is compatible with engaging them in rational persuasion. But not only is a minimal distrust in an agent's reasoning compatible with rational persuasion, it also seems necessary to

motivate rational persuasion more generally. As explained earlier, what distinguishes rational persuasion from rational manipulation is that the former functions to promote the target's rational capacities of the target instead of undermining them. This requires that the persuader make a good faith attempt to assist in furthering or protecting the epistemic interest of the target. But just as a good faith attempt requires at least a minimal trust in the target's epistemic agency, it also requires at least a minimal distrust in the target's epistemic agency. To illustrate, imagine a scenario in which David defers to Rebecca as an epistemic authority on a given topic, and has complete trust in Rebecca's epistemic abilities. In this case, David's motive to intervene cannot be to further or protect Rebecca's epistemic well being since he already has complete trust in her to secure her own epistemic interest. In other words, if the target is viewed as sufficiently competent to look after her own epistemic welfare, the persuader has no reason to advance or protect said welfare.

In sum, a minimal distrust of the target's epistemic agency is necessary for rational persuasion and thus can be charitable when it enables the participant's stance. On the other hand, when a target's rational capacities are sufficiently distrusted so as to preclude rational persuasion, this can constitute taking an objective attitude. As I argued in section 1, the paternalistic worry presents itself when the objective attitude is taken in these latter cases.

Section 3: Rational persuasion as paternalism?

In the previous section, I explained how David should charitably interpret Rebecca's moral position on vaccines. Interpreting her charitably enables him to treat her with the participant stance by engaging her in rational persuasion. However, it might not be entirely clear that David should actually intervene and engage Rebecca in rational persuasion. One might think that it comes off as condescending and belittling to simply begin persuading someone out of their moral beliefs in favor of your own. George Tsai (2014) takes this claim a step further, and argues that the provision of good reasons can actually be paternalistic when they preempt the target's deliberative process of gathering and weighing evidence for themselves. If this is correct, my prescriptive view has the problem of

switching out one paternalistic act for another. In this section, I want to clarify my prescriptive view that David should take the participant stance and rationally persuade Rebecca, and explain how David can do this without treating her paternalistically.

According to Tsai, rational persuasion is paternalistic when three jointly sufficient conditions are satisfied: i) A is motivated by distrust in B's capacity to independently gather and weigh evidence in a manner that is conducive to B's well-being; ii) A conveys this distrust to B either directly or indirectly; iii) A intrudes on B's deliberation, preventing B from the kind of independent engagement with reasons necessary in making a self-determined decision (p. 97). The driving example Tsai uses to illustrate how rational persuasion can be paternalistic is the following case. Peter learns that his daughter, Claire, has just been accepted to a few law schools and philosophy PhD programs.

Peter thinks that Claire is insufficiently capable of carefully considering everything that she ought to be considering before making her decision, and worries that Claire will make the wrong decision. Peter strongly prefers that Claire choose law over philosophy, believes that Claire would be making a serious mistake should she choose the latter, and is not open to Claire persuading him that he is wrong. This evening, Peter presents Claire with lots of information—income statistics, placement data, labor market outlook, testimonials on the climate for women, and so on—about both professions. Peter provides this information with a view to getting Claire to choose law school over philosophy graduate school, making his intentions plain to Claire (pp. 93-94)

Tsai claims that Peter has acted paternalistically because his intervention is *motivated* by his distrust in Claire's ability to rationally deliberate on her situation. Notice that Peter's worry is not simply that Claire might choose philosophy over law, but more precisely, that she will arrive at her decision due to inadequate information and flawed reasoning. Moreover, Tsai thinks the timing of Peter's intervention *conveys* to Claire his distrust, specifically in her ability to gather and evaluate the sort of evidence required to make a well-reasoned decision. After all, Peter knows Claire does not have

to make a decision for another month, and yet he chooses to present a large amount of evidence to her that very evening. Finally, Tsai thinks Peter has acted paternalistically because his intervention *intrudes* on Claire's deliberative process, thereby preventing her from independently engaging with the reasons in a way necessary for her decision—regardless of what she decides—to properly count as hers. That is to say, by coming to Claire with reasons that are meant to persuade her to choose an outcome that *he* desires, Peter prevents Claire from fully self-authoring her decision.

Similarly to Peter, a possible worry for my view is that David's intervention conveys and is motivated by a distrust in the target's ability to independently deliberate on the matter at hand. That is, David rationally persuading Rebecca seems to satisfy jointly sufficient conditions for an act to be paternalistic. Although I agree that David's intervention satisfies all three of the aforementioned conditions, I think the objection is weakened when we understand each condition as existing on a spectrum, ranging from unproblematic cases to the paternalistic. Because of the specific features of Tsai's example, it might seem reasonable to think that Peter treats Claire paternalistically in his attempt to rationally persuade her. However, I argue that Tsai's example of paternalism relies on features extraneous to the method of rational persuasion. When we modify these features, I think Tsai's example illustrates unproblematic instances of *distrust*, *conveyance* of distrust, and *intervention*.

Distrust:

As stated earlier, a minimal distrust in the target's epistemic agency is required in order to motivate the persuader to engage in rational persuasion. However, such distrust need not be objectionably paternalistic when we consider the general epistemic limitations all agents share (e.g. time, attention, and resources). These limitations bring with them a necessary dependency each of us have on other individuals and institutions to assist in our general knowledge creation. As to when distrusting another's epistemic agency is paternalistic or not depends, in part, on whether it turns out to be an appropriate degree of distrust warranted for the epistemic limitations. Paternalistic distrust seems to entail, among other things, both an excessive and unwarranted degree of distrust in the target's epistemic agency. In the case of Peter, both the sheer amount of evidence, as well as the time at

which he approaches her, suggests an intervention geared more to supplanting her rational capacities than one supporting them. Alternatively, an appropriate amount of distrust from Peter would help to assist, rather than replace Claire's deliberative abilities, and thus would allow Claire the opportunity to hopefully make a more informed decision. For example, rather than thinking Claire would wind up making the wrong decision entirely, Tsai's case would be closer to a morally unproblematic intervention if Peter was motivated instead by thinking that Claire might have missed a specific article, statistic, or piece of information in favor of law school.

Convey:

Although the degree to which a person distrusts someone will affect how that distrust is conveyed, there are ways of communicating distrust that are more respectful than others. When assessing Peter's distrust in Claire, it is important to point out that his worry for both his daughter's well-being and her ability to secure it is not, on its own, constitutive of objectionable paternalism. We might even think that, as a parent, Peter's distrust in Claire reflects a somewhat understandable, albeit excessive concern for her during a time in which his daughter faces a major life decision. However, the direct way in which Peter seems to convey his distrust is less excusable. As Tsai notes, the way one presents facts and reasons to another can make the difference between permissible and impermissible instances of rational persuasion, as modes of presentation that incorporate self-effacing indications such as hedging and prefacing can help to convey humility and acknowledgment that the decision is ultimately the target's to make (Tsai, p. 105). A large part of what makes Peter's intervention problematic is that he fails to mask his distrust in this way. The forcefulness and certainty in which he presents his evidence conveys an air of authority that, in turn, conveys to Claire a distrust in her capacity as a deliberative agent. But this is not to say that conveying any amount of distrust is impermissible. Because rational persuasion implies, at the very least, a minimal distrust in the target's ability to independently gather and weigh evidence, this will often be conveyed to the target in some way or another. However, it is *how* this distrust is conveyed that makes the difference between permissible and impermissible cases of rational persuasion.

Intervene:

Turning now to Peter's intervention, I think a significant detail of the case is that Peter, despite having just presented Claire with reasons in favor of law school, is unwilling to be persuaded by reciprocating rational engagement with her reasons. Although Peter treats Claire as rationally competent to recognize reasons, the fact that Peter is unwilling to open himself to Claire's persuasion suggests that he fails to treat her with, what Colin Marshall refers to as, "rationality respect." Drawing from a Kantian notion of respect, Marshall explains that rationality respect "demands that we engage with each other as fellow reasoners and beings deserving of justification...to sincerely offer each other reasons and call out each other's errors, but also to be open to learning from each other" (p. 11). Peter's insistence on unilaterally providing reasons serves as a significant reason for why Peter's intervention is objectionable, as it serves to suggest that he does not see Claire as a rational equal. This is significant because paternalism is in part constituted by a perceived inequality between rational agents (Shiffrin, 2000; Quong 2011). If Peter can intervene in Claire's independent rational deliberation—Tsai's third condition for paternalistic treatment—*without* closing himself off to her reasons, it seems less clear that Peter's intervention would be impermissible given he opens himself up to being persuaded by Claire. Fortunately, opening himself to Claire's persuasion seems entirely consistent with his intervention. Therefore, like with Tsai's first two conditions of paternalistic rational persuasion, the way someone intervenes will matter.

In this section, I have aimed to show that each of Tsai's conditions can be understood on a spectrum of cases ranging from the objectionable to permissible. Although the case of Peter and Claire might seem to constitute objectionable paternalism, I have argued that Tsai's jointly sufficient conditions inadequately support this claim. This is not to say that these conditions are not valuable in identifying objectionable cases of paternalism. Rather, my point is that Tsai's conditions are not jointly sufficient to show why the case of David and Rebecca—as well as cases like it— is objectionably paternalistic. This means that the permissibility of David's intervention of Rebecca, although satisfying

all three of Tsai's conditions, need not constitute a paternalistic intervention given an appropriate example of distrust, conveyance of distrust, and intervention.

Section 4: Unproblematic examples of forgoing rational persuasion

I have argued there is a *prima facie* obligation against forgoing rational persuasion out of distrust for others' capacity to form the correct judgment. In these cases, I have argued that we should charitably interpret our interlocutor so as to take a participant stance and engage them in rational persuasion. However, I do not think a charitable interpretation should *always* be applied in order to enable rationally persuading our interlocutor. As I've mentioned already, it is morally unproblematic to forgo rationally persuading someone when the reasons for doing so have nothing to do with distrust in the target's rational capacities. Nevertheless, it might not always be clear when the reasons for forgoing persuasion are based in distrust or not. There are two types of cases that I think can be particularly difficult to discern from problematic cases. The purpose of this section is to clarify and distinguish these two types of cases from problematic ones. In doing so, I hope to further clarify the obligation against forgoing rational persuasion by specifying its application.

4.1 Cases involving practical considerations

As previously explained in section 2, when our interlocutor expresses a mistaken or irrational belief, it can be charitable to attribute their belief to an error in reasoning rather than a rational incapacity to form the correct judgment on the matter altogether. By interpreting them charitably in this way, we interpret what they say as non-constitutive of their capacity to judge correctly on the matter so as to take the participant stance required to engage with them in rational persuasion. However, sometimes a charitable interpretation will not amount to taking the participant stance and rationally engaging one's interlocutor. That is, sometimes it will be permissible to forgo rationally

persuading someone despite having interpreted them charitably as having the capacity to judge correctly on a given matter.

One such example of when forgoing rational persuasion is permissible is when the reason for doing so is out of acknowledgment that there exist fundamental differences in worldviews that cannot be reconciled. The reason to forgo persuading someone here doesn't have to be out of distrust in their rational capacities, but rather it can be out of acknowledgment that some worldviews are simply incommensurable with others. Maybe you do believe your interlocutor is rationally capable of understanding your arguments, but she will still choose her way of seeing the world over yours due to fundamentally different belief systems. On the view I have presented, whether or not forgoing persuasion is paternalistic depends on if the motive is grounded in distrust for the *rational capacities* of the target. If the persuader attributes the target's unwillingness to adopt a change in belief to the fact there exists incommensurable worldviews, then forgoing persuasion is not paternalistic. Of course, the persuader's assessment of why certain beliefs are incommensurable with others will make a difference. For example, if a believer of the Catholic faith thinks his religion is incommensurable with Islam because he thinks Muslims are deeply confused in their reasoning and incapable of arriving at the truth, then his motivation seems to be grounded in a distrust for the target's rational capacities. On the other hand, if the Catholic believer thinks the incommensurability between religions is not indicative of the rational capability of adherents, then his decision to forgo persuasion does not seem paternalistic. While there is a fine line here between cases that are respectful and cases that are paternalistic, the primary difference lies in the motive behind forgoing persuasion, and whether that motive lies in distrusting the rational capacities of the target.

Another example of when forgoing rational persuasion is permissible is when the reason for doing so has to do with the persuader having insufficient information or evidence about a topic or area of discussion. Unlike with David, there will be times when people fail to possess the information or evidence needed to make an effective attempt at persuading their interlocutor. Take the following case of Simon who, while confident that his government has a moral obligation to mitigate the effects and the increasing threat climate change poses to future generations, is unsure of how to go about

persuading his friend Natalia of this view. Simon's confidence in this view comes from his general trust in science. Apart from overhearing conversations from friends and family on the topic, he is largely unfamiliar with both the ethical arguments and empirical evidence pertaining to the moral obligations surrounding climate change. In this case, there is no worry of paternalism if Simon decides to forgo rationally persuading Natalia if his reason for doing so is a general distrust in his own ability to present her with clear and reliable information.

Moreover, even when someone does have the requisite evidence and information at their disposal, it may still be difficult to argue, let alone argue effectively, with someone who they are not well acquainted with. This takes a certain level of confidence and charisma that not everyone possesses. This challenge is amplified when the arguments pertain to hot button moral issues such as vaccinations and climate change. It takes confidence and a fairly extroverted personality for someone with a non-combative personality to act as David does and challenge a person they have just met. When the reasons for forgoing rational persuasion have to do with feeling the general uncomfortability that comes with these challenging interactions,—rather than a distrust of the target's rational judgment—I think doing so is entirely unproblematic.

4.2 Cases involving respect-based considerations

If you have ever debated hot button moral issues, you can understand that it can be taxing to engage with people with conflicting beliefs. This point is especially true when we acknowledge that engaging in ethical rational deliberation with someone often brings with it cognitive and epistemic demands of its participants, such as exercising open mindedness and mutual respect (Breakey, 2019). But what about rationally persuading people who hold repugnant beliefs? The difference with these views is that they are typically hostile and potentially harmful to those that oppose them. I think cases like these pose a deeper concern for the persuader's general well-being and self-respect. It's one thing to disagree with someone whose view is partially misaligned with your own, but quite another when it's a view that fundamentally opposes your worldview

As I have argued, the participant stance is enabled when David takes a charitable interpretation of Rebecca's view on vaccines. However, charitably interpreting an interlocutor's opposing position will often ask of the persuader to extend a deeper sense of empathy and understanding toward a repugnant view than one is typically accustomed to doing. One's comfortability in doing this may vary, though in some cases it may be deeply troubling for the persuader to extend empathy and understanding to views that fundamentally conflict with their own. An illustrative example, albeit an extreme one, is from the movie *Manhunter* (1986), in which special agent Will Graham is tasked with apprehending a serial killer by the name of the Tooth Fairy. In his search for the killer, Graham employs his uncanny ability to enter the minds of his targets by inhabiting the perspective of a psychopath. Although making him incredibly effective at what he does, taking the perspective of his targets is taxing on Graham's psychological well-being, as his identity as a father, husband, and overall law-abiding citizen is threatened by the polar opposite values that he must inhabit in order to be effective at his job. For example, as a father, Graham is particularly disturbed by identifying with the Tooth Fairy's fractured sense of family, and empathizing with the dark impulses that stem from his upbringing.

The concern raised by the *Manhunter* example is that when it comes to extending a charitable interpretation to abhorrent beliefs, we risk opening ourselves up to perspectives that might be too much for us to handle. The effect from this can sometimes be positive, as one is allowed to inhabit new perspectives. Moreover, by engaging repugnant views head-on through rational deliberation, one can provide the epistemic friction that can serve to alleviate bias and misinformation (Ferkany, 2021). Other times, however, an agent can feel alienated from one's deepest commitments and projects. But the point is not that we should forgo extending charity to views that simply differ from our own. Rather, it is that we should be weary of the psychic pain that may follow from inhabiting the perspective of someone whose views are hostile to one's own fundamental commitments. Of course, not all cases will be as straightforward as the decision to forgo persuading the psychopath. Nonetheless, I think it should be left to the discretion of the persuading party whether or not to extend charity to views that pose a threat to their own worldview. We should not demand from people that they extend charity in these situations if it puts them in positions of psychic harm.

I also think that the persuader should avoid putting the *target* in a position that risks undue harm.⁹ For example, let us modify our starting case and imagine that the diner that day is packed with customers that have just arrived from watching a football game, requiring Rebecca to wait on six different tables of rowdy and hungry fans. In this revised version of the case, it seems likely that it would be a serious inconvenience for Rebecca to engage with David's arguments, as it would affect her duties as a waitress. In this case, if David correctly identifies the situation and forgoes persuading Rebecca out of basic consideration for her responsibilities, he would not treat her paternalistically. While I think we should not forgo persuading our interlocutor out of distrust for their capacity to engage with our moral reasons, there are certainly other reasons for why we might forgo persuading our interlocutor that are unproblematic, and sometimes even praiseworthy. Even if we take a case where the diner is empty, David might nonetheless owe it to Rebecca not to attempt persuading her of his position on vaccines in her place of business. If he does forgo persuading her out of respect for her work, and not out of distrust for her rational capacities, then he does not treat her paternalistically. On the other hand, it would seem unobjectionable for David to persuade her if she decided to attend his lecture. Putting aside cases of distrust, my point is that I think there can be a time and place for rational persuasion. In these scenarios, the persuader will have to be attentive to the context in which he chooses to engage in persuading his interlocutor, since there is a responsibility to guard against causing the target undue harm.

Conclusion

It is no secret that deep moral and political disagreements are rarely resolved solely through rational persuasion. However, there are times when forgoing rational engagement with our interlocutor treats them disrespectfully. Namely, we treat our interlocutor paternalistically when we forgo persuading them out of distrust for their capacity to judge correctly. In this paper, I have argued there is a *prima facie* obligation against forgoing rational persuasion out of distrust for others' capacity

⁹ Although this might include psychic harm, I focus on harm more generally here.

to form the correct judgment. Rather than distrusting our interlocutor's capacities, I have argued that we should charitably interpret our interlocutor so as to treat them as participants capable of engaging in rational persuasion.

References

- Begon, J. (2016). Paternalism. *Analysis* 76(3), 355-373.
<https://doi.org/10.1093/analys/anw040>
- Breakey, H. (2023). The ethics of arguing. *Inquiry*, 66(4), 589–613.
<https://doi.org/10.1080/0020174X.2019.1637776>
- Ferkany, M. (2021). How and Why We Should Argue with Angry Uncle: A Defense of Fact Dumping and Consistency Checking. *Social Epistemology*, 35(5), 533–545.
<https://doi.org/10.1080/02691728.2021.1930275>
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Pantheon/Random House.
- Kahan, D. and Braman, D. (2006) Cultural Cognition and Public Policy. *Yale Law & Policy Review*, 24(87) 147-171.
<https://ssrn.com/abstract=746508>
- Kennett, J., Fine, C. (2009) Will the Real Moral Judgment Please Stand Up? *Ethical Theory and Moral Practice* 12(1), 77–96.
<https://doi.org/10.1007/s10677-008-9136-4>
- Mann, M. (1986). *Manhunter* [Film]. De Laurentiis Entertainment Group.
- Marshall, C. (2024). Political persuasion is prima facie disrespectful. *Journal of Moral Philosophy*, 22(5-6), 584-617.
<https://doi.org/10.1163/17455243-21050017>
- McGrath, S. (2022) Please Keep Your Charity out of My Agency: Paternalism and the Participant Stance. In *Oxford Studies in Normative Ethics (Vol. 12)*. Oxford University Press.
- Price, T. L. (2020). *Leadership and the ethics of influence*. Routledge.
<https://doi.org/10.4324/9780429449161>
- Quong, J. (2011). *Liberalism Without Perfection*. Oxford: Oxford University Press.
- Rini, R. (2018). Abortion, Ultrasound, and Moral Persuasion. *Philosophers' Imprint*, 18(6): 1–20.

<http://hdl.handle.net/2027/spo.3521354.0018.006>

Shiffrin, S. V. (2000). Paternalism, Unconscionability Doctrine, and Accommodation.

Philosophy & Public Affairs, 29(3), 205–250.

<https://doi.org/10.1111/j.1088-4963.2000.00205.x>

Strawson, P. F. (1963). Freedom and resentment. *Proceedings of the British Academy*, 48(), 187–211.

Schroeder, M. (2019) 'Persons as Things', in Mark Timmons (ed.), *Oxford Studies in Normative Ethics*, 9(): 95-115.

<https://doi.org/10.1093/oso/9780198846253.003.0005>

Tsai, G. (2014). Rational Persuasion as Paternalism. *Philosophy & Public Affairs*, 42(1): 78-112.

<https://doi.org/10.1111/papa.12026>

Persuasion and Deference

Introduction

To be rationally autonomous is to make your own decisions and come to your own conclusions, free from external pressure or undue influence. Types of influence tactics that typically pose a threat to rational autonomy are manipulative and/or coercive. Rational persuasion on the other hand, is typically thought to enable rational autonomy by providing the target with reasons, facts, and evidence that encourage rational deliberation, rather than undermine it. However, recent attention has been drawn to the ways in which the mere provision of reasons can pose a threat to rational autonomy (Tsai 2014; Price 2020). For example, Terry Price argues that despite its appeal to the rational capacities of the target, rational persuasion can serve as a tactic that *gets* the target to act and think in a way the persuader chooses, rather than *letting* them decide for themselves. As Price points out, when we attempt to rationally persuade our target, we do not simply aim to give them reasons to deliberate over, but do so with the aim of getting them to accept our view (p. 65). The aim of rational persuasion can be contrasted with the aim of teaching, where the aim of the latter is to provide students an opportunity to learn without trying to get them to accept any specific conclusion (Freire, 1970). Employing rational persuasion to get students to adopt a specific conclusion can be problematic because it can impede the students' ability to think for themselves. The same goes for rational persuasion more generally, where getting others to think or do what *we* decide *they* should think or do runs the risk of precluding them from exercising their rational autonomy when it gets them to defer to our judgment (p. 74).

I argue that while attempting rational persuasion can sometimes result in the target deferring to the persuader's judgment, it can be morally problematic not because it risks precluding the target from exercising rational autonomy, but rather because it can fail to reciprocate respect for the target's rational agency. While all cases of precluding the target's rational autonomy are cases that also fail to respect the target's rational agency, not all cases of failing to respect the target's rational agency are ones

that preclude the target's rational autonomy. I argue that while rational persuasion depends on the target exercising her rational autonomy by expecting her to rationally appreciate reasons, the persuader's attempt can nonetheless undermine the target's status as a rational agent by failing to treat them as a rational equal. This unequal treatment of the target is expressed when the persuader presents reasons unilaterally on a subject in which they have no justifiable expertise. By unilaterally providing reasons to the target, the persuader signals to the target they are unwilling to engage with any of the target's counter-reasons. The provision of unilateral reasons gives off what I refer to as an 'authoritative air.' In response to Price, I will argue that rational persuasion, by definition, cannot preclude the target's rational autonomy since it requires the target to rationally appreciate the reasons given them. However, a persuader *can* fail to reciprocate respect for the target's rational agency when the persuader is unwilling to engage with the target's counter reasons. The type of respect the persuader can fail to show the target's rational agency is what Colin Marshall (2024) calls "rationality respect," which "demands that we engage with each other as fellow reasoners and beings deserving of justification...to sincerely offer each other reasons and call out each other's errors, but also to be open to learning from each other" (p. 11). Fundamentally, treating someone as a fellow reasoner requires that we treat them as sufficiently competent to engage with our reasons as well as give reasons of their own. When reasons are given unilaterally, the persuader fails to treat the target as a fellow reasoner by treating her merely as a *receiver* of reasons, and not also as a *giver* of reasons.

My argument proceeds as follows: In section 1, I distinguish between problematic and unproblematic cases of deference. I then explain why deferring to what I refer to as, 'identity-constituting beliefs', is problematic because it can fail to fulfill an important ideal of rational agency: self-authorship. In section 2, I consider two ways in which rational persuasion might elicit deference from the target. I argue that both considerations inadequately account for the motive behind rational persuasion, properly understood. The problem with rational persuasion in these cases is not that it risks precluding the target's rational autonomy, but rather that the attempt can express an authoritative air by failing to reciprocate rationality respect. In section 3, I provide two explanations for

how Price can be right about the risk of the target deferring in cases of rational persuasion, but just mistaken about rational persuasion as the cause of the target's deference.

1. The moral status of deference

In order to make sense of Price's claim that rational persuasion can problematically elicit deference from the target, it will be helpful to first distinguish between problematic and unproblematic cases of deference. Let us begin with cases in which it is unproblematic to defer to the rational judgment of another. For example, deferring to scientific experts is typically thought to be necessary in knowledge creation given our finite epistemic resources such as time, attention, and information. An unproblematic case would be for me to defer to a botanist—who has devoted their life to studying the molecular structure of plants—on what a plant's molecular response to heavy metal contamination would be, since they are positioned as a more reliable source of true beliefs regarding such matters than myself as a non-specialist. That many of us defer to experts like the botanist for our beliefs about the natural world is not only generally unproblematic, but also seems necessary in order to hold a range of true beliefs. If we imagine the ultimate autonomous agent who only forms beliefs based on independent empirical research, it would simply be impossible for them to have the number of beliefs they otherwise would have from depending on the knowledge of others. The necessity and justification of outsourcing certain beliefs applies to non-scientific beliefs as well. For example, imagine you find yourself lost while driving on some country back road without cell phone reception. Rather than rely on your own sense of direction, you stop at a local gas station to ask the gas attendant for directions back to the main road. In this case, it would be entirely unproblematic for you to defer to the gas attendant's sense of direction and to form the belief, per the instructions given, that the main road is two miles north. Similar to when we defer to scientists for beliefs about the natural world, it is unproblematic to defer to the local gas attendant's sense of direction in this case because, as a local, he appears to be better positioned than yourself to hold true beliefs about the location of the main road.

However, deference may seem problematic in other matters. In particular, philosophers are skeptical about the permissibility of deferring to moral beliefs. In the literature, this is primarily because of doubt as to whether there can be moral experts in the same way scientists are experts in scientific matters (McGrath, 2019). But I will not broach the question of moral expertise here. Instead, I want to look at what I will refer to as, *identity-constituting beliefs*, defined as beliefs that bear on areas in our lives that are generally thought to be especially influential in shaping who we are (e.g. moral matters, family affairs, career decisions, etc).¹⁰ As stated at the outset, I am concerned with how deference can undermine the autonomy of an agent. While the question of moral expertise is relevant to evaluating moral deference, I focus on identity-constituting beliefs because deferring to them can help explain the problem with the target not having agency over her moral beliefs, regardless of her interlocutor's expertise. If I am correct in broadening the scope of concern outside strictly just moral deference, this will be significant for Price's argument insofar as this would imply a broader scope of potentially impermissible cases of rational persuasion.

To help explain the problem with deferring to others on identity-constituting beliefs, let us first look at a distinction Sarah McGrath gives us between *transparent* and *opaque* evidence for belief. Transparent evidence in support of a belief suggests that a proposition is true while also providing information for *why* that proposition is true. On the other hand, opaque evidence for a proposition only suggests that the proposition is true without explanation for *why* it is true (pp. 89-90). To illustrate, take the case given by McGrath, where a math test contains a question that asks you to prove a mathematical proposition. You think to yourself it unlikely the test would ask you to prove an incorrect proposition, and thus take this as evidence for why the proposition is true. In this case, the evidence you have is opaque because it only suggests the truth of the proposition without explaining why the mathematical proposition is true. In contrast, if you were to provide a detailed, clear and accurate proof of the proposition, this would count as transparent evidence since it not only suggests its truth, but also explains why the proposition is true (p. 90).

¹⁰ See Van Leeuwen (2023) for a similar use of the term.

With this distinction in mind, I think that deferring identity-constituting beliefs can fail to fulfill an important ideal of rational agency: self-authorship. Drawing from the literature, let us define self-authorship as the act of modeling, deliberating, and reflecting on the reasons for the actions, beliefs, and values most central to one's identity (Dover 2023; Garnett 2023; Tsai 2014; Paul 2014). Insofar as one only has opaque evidence for a belief, one will be prevented from modeling, deliberating, and reflecting on the reasons in support of that belief. But why is it important to self-author our beliefs? After all, as we've seen, it seems unproblematic to defer to scientists about many of our beliefs about the natural world, so why do we need to self-author our lives here? I think it matters for identity-constituting beliefs—unlike with scientific beliefs—that you believe them for reasons that you can model, deliberate, and reflect on. This is especially true when the actions that flow from these beliefs are, in turn, not self-authored. To help motivate this idea, consider the following example: Jillian is a senior in college applying to grad school. Jillian's dad has always believed that Jillian should go to med school to become a doctor. Since she became aware of her father's belief, she has never really thought about or questioned her father's judgment on the matter, and believes too that she should go to med school, but for no other reason than because her father believes she should go. In other words, Jillian defers to her father's judgment on the matter by having only opaque evidence for her belief. In order to avoid ambiguity, let us stipulate that Jillian does not believe she should attend med school for any significant second-order reasons, such as wanting to make her dad proud. If Jillian does have a second-order reason for believing she should attend med school, it will be an insignificant one (e.g. Jillian did not feel like thinking about college at the time, and so adopted her father's belief for the sake of convenience).

I think intuitively there is something fishy about the above case, since one's beliefs and corresponding decisions regarding education and a career path are important to one's own life in a way that clearly is not for beliefs and corresponding decisions regarding, say, the weather. Although of course we can't be expected to self-author all of our decisions, beliefs, and actions, some philosophers think there is, what Daniela Dover calls, a "heightened scrutiny" when it comes to interpersonal influence over areas of our lives commonly thought to belong to our individual sphere of control

(Dover, pp. 1-2). The main reason for a heightened scrutiny is, as Dover points out, concern for one's opportunity to self-author significant parts of their life.¹¹ George Tsai makes a similar point when he argues for the importance of allowing others to, at least preliminarily, "work things out for themselves", where one has the opportunity to "recognize and weigh" the reasons behind important decisions and beliefs in their lives. Tsai refers to the questions that one asks oneself when deliberating about such decisions and beliefs as 'eudemonistic' questions (p. 100). Questions of this sort are distinctively personal, examples which include 'what sort of life do I want to lead?' and 'what is my purpose in the world?' Answering these questions for ourselves—without the influence from others—is an important part of a self-authored life. When we are not given the epistemic space for this "purer" and "more direct" deliberative authority over these kinds of questions, "our decisions can feel, indeed be, less fully our own" (p. 100).¹² Insofar as we value the opportunity to self-author personally significant parts of our lives, we have an obligation to respect other agents' deliberative space when deciding intuitively personal questions. This point seems especially true since, unlike with scientific matters where it is often the case that others are more qualified on the topic, we typically are thought to have authority over matters concerning our personal lives. That we are typically thought to have authority over such matters helps to explain what, I claim, appears to us as 'fishy' with Jillian deferring to her dad's belief that she should attend med school. Whether or not Jillian's dad is blameworthy for Jillian's deference will depend on the details of how he approached and communicated his belief to Jillian. More specifically, evaluation of Jillian's dad will depend on whether he promoted or detracted from Jillian's ability to engage independently with the reasons for her belief. In the next section, I will talk about how an influencer's intentions can help distinguish permissible and problematic cases of deference. For now, I hope to have shown that deferring on matters intuitively thought to be personally significant (i.e. identity-constituting beliefs) can risk preventing us from self-authoring our lives.

2. Rational persuasion and deference: Two explanations

¹¹ Dover ultimately rejects the 'heightened scrutiny' view.

¹² See Dover for an informative, albeit different, summary of how Tsai aligns with the heightened scrutiny view.

In this section, I look at the role that rational persuasion plays in cases of problematic deference, focusing exclusively on identity-constituting beliefs. I will henceforth use the term ‘rational persuasion’ to refer to the provision of reasons communicated to the target with the intention of changing the target’s mind about some identity-constituting belief.¹³

It might strike us as peculiar as to how the mere provision of reasons can elicit problematic deference from the target, as Price argues. After all, rational persuasion is typically lauded by philosophers as the ‘gold standard’ precisely because it appeals to the rational capacities of the target. That is, the idea that rational persuasion *promotes* the rational capacities of the target seems to be in conflict with our intuitions about problematic cases of deference, which tell us that deferring to another entails that the target *outsources* their deliberative activities, rather than work it out for themselves. I think an initially attractive explanation is rational persuasion elicits deference in what I refer to as, ‘blunt deference’: A risks getting B to defer when A tells B, "you should believe X because of Y," *without* an expectation that B rationally appreciates why Y supports X. To help illustrate how this formulation captures deferential treatment, consider the following example:

Blunt deference: Agnes wants to persuade her daughter, Olivia, that it is morally wrong to ever consume animal products. Agnes presents Olivia with a large amount of scientific evidence on why abstaining from animal products helps minimize animal suffering and harm to the environment. Agnes doubts that her daughter will understand why exactly the evidence supports the moral obligation to ever abstain from animal products (as opposed to, say, merely reducing consumption of animal products), but hopes that it will get Olivia to come around and adopt a strictly vegan lifestyle. After all, Agnes doesn’t particularly care if Olivia understands why exactly the evidence supports the wrongness of ever consuming animal products, as long as she ends up sharing her mother’s belief in the wrongness of ever consuming animal products.

¹³ As opposed to some scientific belief, for example.

This may strike us as an initially attractive explanation for how rational persuasion can elicit problematic deference, since the persuader risks getting the target to defer to X when the reasons in support of X are not expected to be rationally appreciated. In other words, offering reasons that are not expected to be rationally appreciated risks precluding the target from exercising their rational autonomy because it enables the target to adopt beliefs based on opaque evidence. Although initially appealing, I think we should reject this solution to the apparent conflict between rational persuasion and problematic deference, since Agnes's expectation that her daughter will not rationally appreciate her reasons is at odds with the aim of rational persuasion, properly understood. This is because when an agent decides to rationally persuade another agent, rather than manipulate or coerce them, it matters to the persuader that the target genuinely engages with the reasons, and open themselves to a change of *mind*, rather than simply a change of *behavior* (Rini, pp. 4-5). This is what separates rational persuasion from other methods like manipulation and coercion. Where manipulation and coercion function to undermine or subvert the target's rational capacities in order to achieve the intended change of behavior, rational persuasion aims to promote the target's rational capacities by providing reasons that the target is expected to rationally deliberate and reflect on (Rini, pp. 5-6). This expectation of the target implies that the target is expected to understand *how* the reasons support the belief, thus requiring the target to have transparent rather than opaque evidence. In other words, when attempting to rationally persuade others of identity-constituting beliefs, it matters to the persuader that the target engages with the reasons provided, such that they rationally appreciate why X supports Y.¹⁴

That Agnes does not expect Olivia to rationally appreciate how the reasons she presents support her claim is more characteristic of the motive behind something like 'rational manipulation,' where "reasons are provided but their point or purpose is not to advance the other's rational reflection or change of mind. Instead of sustaining or promoting the rational process, the aim in rational

¹⁴ Contrast this with the mere provision of reasons outside the context of moral and identity-constituting beliefs, where this is not necessarily the case. For example, when someone asks us for directions it will typically not matter to us that the person changes their mind in light of engaging with the reasons we have provided, but rather only that they arrive at their desired location. In these types of cases, it is enough that they simply follow our directions, even if they are not fully convinced that we are correct.

manipulation is to subvert it” (Tsai, p. 90). The motive behind Agnes’s provision of reasons is characteristic of the motive behind rational manipulation, where the reasons given are not intended to advance the target’s rational deliberation, but rather are intended to bypass the rational capacities in an effort to influence the target. Although ‘intended’ might be too strong of a description in this case, Agnes’s expectations that her daughter won’t understand how the reasons support the belief is, at the very least, indicative of her indifference to promoting her daughter’s rational capacities. As Tsai suggests, it is the motive behind the provision of reasons that constitutes the difference between rational persuasion and rational manipulation.¹⁵ Since Agnes’s efforts do not involve the requisite motive necessary to count as rational persuasion, but instead, are characteristic of that in rational manipulation, we must conclude that it is not rational persuasion, properly understood, that elicits deference from Olivia.

An alternative way to make sense of how rational persuasion might risk getting the target to defer is suggested in the following case of, what I will refer to as, ‘subtle deference’: A tells B, “you should believe X because of Y,” *with* an expectation that B rationally appreciates why Y supports X, but *without* leaving any room for B to argue back (A might say, “look, stop arguing and just think about what I’ve said - you’ll understand why I’m right”). Now consider a revised case of Agnes that meets this description of deference:

Subtle deference: Let us imagine that it matters to Agnes that Olivia becomes a vegan because it helps the environment and prevents animal suffering. That is, it matters to Agnes that Olivia rationally appreciates the reasons for adopting the belief that it is immoral to consume animal products. If Olivia decides she doesn’t care about animal suffering and helping the environment, it is not enough for Agnes that Olivia becomes vegan simply to appease her mother. Although Agnes is open to helping Olivia understand how the evidence supports the belief that one should abstain from animal products, she is not open to hearing any counter

¹⁵ See Breakey (2023) for a similar view of rational manipulation.

arguments from Olivia. Agnes is confident in her moral position on the matter, and thinks that addressing Olivia's opposing viewpoints is a waste of time.

The first thing to observe is that, unlike with blunt deference, subtle deference is compatible with rational persuasion since the persuader is expected to appreciate how X supports Y. When the target is expected to appreciate how X supports Y, the target must be put in a position to reflectively endorse X with transparent evidence, rather than defer to X with only opaque evidence. In other words, the motive behind subtle deference does not risk the impediment of the target's rational deliberation in the way blunt deference does, since the target is encouraged to decide for themselves through rational engagement with the reasons in support of the belief.

Even though the target is asked to use their rational autonomy in a way they are not in blunt deference, there still seems to be something wrong with Agnes's refusal to listen to any of Olivia's arguments. I think subtle deference is problematic because it expresses a dismissive and superior attitude characteristic of deferential treatment that I will refer to as an 'authoritative air.' The message sent by the persuader's unilateral provision of reasons is something like "decide for yourself, but you will see that I'm right." This message expresses authority in the persuader's self-confidence that they are right and have the answers, but also that the target is incapable of offering reasons deserving of serious consideration.

Moreover, the problem with the authoritative air is especially pronounced given the subject matter of identity-constituting beliefs, since the target is typically considered to command some authority over her own identity-constituting beliefs. As mentioned before, this is in part due to the heightened scrutiny about whether there can be experts on such matters. Again, we can contrast this with scientific beliefs, where expertise gives the scientific community justified authority over questions concerning the natural world. There are other contexts outside of the scientific community that seem to rightly allow for an authoritative air. For example, in contexts where the subject matter falls within the mutually agreed upon power dynamics, such as that involving employer and employee (e.g. beliefs about how best to run a company), the authoritative air also seems permissible. However, when it

comes to one's identity constituting beliefs, I think we should be more suspicious about outside authority than we are about beliefs in other areas of our lives.

To further explain the insult conveyed by the authoritative air, let us invoke the concept Colin Marshall (2024) calls "rationality respect," which "demands that we engage with each other as fellow reasoners and beings deserving of justification...to sincerely offer each other reasons and call out each other's errors, but also to be open to learning from each other" (p. 11). In one sense, Agnes rationally respects Olivia as a fellow reasoner when she offers her evidence to rationally deliberate over (Rini, 2018). However, to be treated equally as a fellow reasoner, Agnes must also be open to hearing her daughter's reasons, given her expectation of her daughter to genuinely engage with her own. Although Olivia is treated rightly as a competent *receiver* of reasons, Agnes's unwillingness to engage with her daughter's reasons prevents her from treating Olivia as a rational equal. It is the lack of reciprocating respect for Olivia's ability to *give* reasons that constitutes the authoritative air. To further illustrate how this lack of reciprocity can be disrespectful, consider the intuitively problematic case of the 'fanatic', who as Regina Rini describes for us is "a person so committed to one particular moral ideal that he will not even consider other moral reasons that might come into tension with it" (pp. 6-7). What explains the problematic behavior of the fanatic in cases of rational persuasion is not simply his unwillingness to engage with our reasons, but it is that his refusal is paired with an expected willingness from others to engage with *his* reasons and arguments. It is this unwillingness to reciprocate rationality respect, and the authoritative air it expresses, that constitutes the problem with subtle deference.

It is worth emphasizing that I think the worry with not reciprocating rationality respect is different from the worry that the target will defer without exercising their rational autonomy. That is, if the target does defer in cases of rational persuasion, it does not follow that the persuader is responsible for precluding the target's rational autonomy even if he closes himself off to the target's counter arguments. This is because, in offering the target reasons he or she is expected to appreciate as evidence, the persuader does not preclude the rational autonomy of the target. The inverse is equally true, where if the target does not defer in cases of rational persuasion, it does make the persuader's closing themselves off to the target's counter arguments any less objectionable. What makes subtle

deference objectionable is not that it risks precluding the target from exercising rational autonomy, but rather because the attempt by the persuader expresses an authoritative air and fails to reciprocate rationality respect for the target.

An upshot of my analysis is that it explains the problem with rational persuasion in terms of a failure to reciprocate rationality respect, rather than the peculiar idea of rational persuasion precluding rational autonomy. I think my account better captures what intuitively can go wrong in cases of rational persuasion where, even though the persuader cares about appealing to the rational capacities of the target, doing so can nonetheless express an authoritative air when done poorly and without proper consideration for the target. Although the persuader encourages the target to decide for themselves, the persuader can express through a unilateral provision of reasons that it's not really a decision at all, and that the correct belief is simply staring the target in the face. Therefore, my account is able to accommodate the intuition that rational persuasion is different from methods like manipulation, as well as the intuition that there can be something objectionable about the unilateral provision of reasons.

3. Reallocating blame for deference

So far I have argued that rational persuasion does not preclude the rational autonomy of the target, and therefore does not elicit the kind of deference Price thinks it does. But a problem still remains in cases where the target—despite the persuader expecting the target to rationally appreciate the reasons given to her— defers to the persuader without rationally appreciating the reasons. That is, something or someone seems to be responsible for undermining the target's rational autonomy when the target defers in these cases, Price would say. But if it is not the fault of rational persuasion, as I have argued, then what is it that undermines rational autonomy in cases of deference where the target adopts a persuader's belief without rationally appreciating why the reasons support the belief? In this section, I want to show how Price can be right about the risk of the target's rational autonomy being undermined in cases of rational persuasion, but just mistaken about rational persuasion as causing this

risk. By explaining what causes deference in these cases, I hope to further explain that if the target does defer without rationally appreciating reasons, it is not the fault of the rational persuader, as Price claims. In the last section I argued for why rational persuasion was not responsible for precluding the rational capacities of the target. This section helps to expand on that point by identifying the source of responsibility in such cases.

What should we make of the case where Olivia defers to Agnes in subtle deference? That is, despite Agnes's expectation that Olivia appreciates how the evidence supports the belief in question, Olivia still defers to the belief without this appreciation of the supporting evidence. In this case, Olivia is left worse off for the reasons I outlined in section 1. But where can we locate the cause of this harm to Olivia? In cases like this, I think it is helpful to distinguish between deference and *efforts* that get the target to defer, where in this case you have the former without the latter. This is a case of inadvertent deference since, as I have argued, it is not Agnes's intention that Olivia's rational autonomy is precluded. But nor does it seem to be Olivia who is responsible for undermining her own autonomy, since she does not possess a second-order reason, and thus does not *choose* to undermine her own autonomy. Therefore, because there is no deferential intention from either Agnes or Olivia, cases like this seem best described as inadvertent deference.

But what about cases where the target *does* choose to defer with clearly defined second-order reasons? Take what Michael Garnett (2023) refers to as cases of 'practical deference,' where "*B* has a disposition of practical deference just in case *B* is disposed to take *A*'s wanting it to be the case that *p* to be an especially weighty reason in itself for bringing it about that *p*, independently of the content of *p*" (p. 198).¹⁶ Applying this formulation to the case of Agnes, let us imagine that Olivia consistently defers to her mom on moral beliefs—despite her mom's efforts to explain why the evidence supports the belief—solely for the reason that her mom wants her (Olivia) to share her moral beliefs. These cases seem problematic given that Olivia fails to exercise her rational autonomy, and routinely defers to what she thinks her mom wants her to believe. As Garnett puts it, "[Agnes's] preferences simply are

¹⁶ Contrast this disposition with the disposition of love and care, where a person is "disposed to prioritize another person's interests." The difference with the disposition of practical deference is that the deferential party is "disposed to do whatever another person merely *wants*" (pp. 204-205). As Garnett puts the distinction, the problem with the disposition of practical deference "is not that she cares about another's good, but that she renders herself subservient to another's will" (p. 204).

[Olivia's] reasons" (p. 198). But it's not clear in these cases that the target's autonomy is undermined without external undue pressure. This is because we typically think the target exercises autonomy in the act of choosing to defer, and thus in order for the choice to be deferential, it must have been unduly influenced. Although we typically look for external explanations of undue influence in these cases, Garnett argues that we can explain the undermining of the target's autonomy without appealing to explanations of external undue influence. This is because the decision to submit oneself to the will of another—where Olivia's reasons are simply her mom's preferences—inherently lacks first-order autonomy, and thus is not self-authored, but rather "ghostwritten" (p. 213). As Garnett explains, self-authorship requires both first and second-order autonomy. When Olivia gives up her first-order autonomy because of deferential second-order reasons, this is analogous to a ghostwritten life, in that the beliefs Olivia defers to are, in a sense, ghostwritten by her, but "dictated by someone else" (p. 213). The analogy helps explain the sense in which we see Olivia as exercising autonomous choice, while also explaining the sense in which we see Olivia as non-autonomous when she chooses to submit to the will of her mother on moral matters.

In this section, I have outlined two ways in which the target's autonomy can be undermined without external undue pressure. One way is when the target *inadvertently* adopts the persuader's moral position (i.e. without second-order reasons) without rationally appreciating the reasons in support of the position. The second way is when the target *chooses* to adopt the persuader's moral position (i.e. with clearly defined second-order reasons) without rationally appreciating the reasons in support of the position. My aim in this section was to show how Price can be right about the risk of the target's rational autonomy being undermined in cases of rational persuasion, but just mistaken about rational persuasion as the cause of the target's deference. By identifying what causes the target to defer in these cases, I hope to have furthered my argument for why rational persuasion does not pose a threat to the target's rational autonomy.

Conclusion

I have argued that the aims of rational persuasion are incompatible with the aim of eliciting deference. Moreover, I have argued that the attempt at rational persuasion can only result in what I have termed, 'subtle deference,' where what is problematic is not the precluding of the target's rational autonomy, but rather a failure to reciprocate respect for the target's rational agency. By implication, my conclusion defends rational persuasion as a respectful method of influence in cases where the persuader willingly opens themselves to the target's reasons.

References

- Breakey, H. (2023). Is Rational Manipulation Permissible? *Ethical Theory and Moral Practice*, 26(2): 199-216. <https://doi.org/10.1007/s10677-023-10382-4>
- Dover, D. (2023). Identity and Influence. *Synthese*, 202(5):1-24. <https://doi.org/10.1007/s11229-023-04279-z>
- Garnett, M. (2022). Ghost-Written Lives: Autonomy, Deference, and Self-Authorship. *Ethics*, 133(2):189–215. <https://doi.org/10.1086/722126>
- Freire, P. (1970). *Pedagogy of the Oppressed* (M. B. Ramos, Tran.). Herder and Herder.
- Marshall, C. (2024). Political Persuasion is Prima Facie Disrespectful. *Journal of Moral Philosophy*, 21(5):1-34. <https://doi.org/10.1163/17455243-21050017>
- McGrath, S. (2019). *Moral Knowledge*. Oxford University Press.
- Price, T. (2020). *Leadership and the Ethics of Influence*. Routledge, Taylor & Francis Group. <https://doi.org/10.4324/9780429449161>
- Rini, R. (2018). Abortion, Ultrasound, and Moral Persuasion. *Philosophers' Imprint*, 18(6): 1–20. <http://hdl.handle.net/2027/spo.3521354.0018.006>
- Tsai, G. (2014). Rational Persuasion as Paternalism. *Philosophy & Public Affairs*, 42(1): 78-112. <https://doi.org/10.1111/papa.12026>
- Van Leeuwen, N. (2023). *Religion as Make-Believe : A Theory of Belief, Imagination, and Group Identity*. Harvard University Press.

“But I Said I’m Sorry”: The Role of Persuasion in the Ethics of Apologies

Introduction

A reparative duty is a duty to repair the break in the treatment of someone worthy of basic respect and consideration. Duties to apologize can be reparative, as an apology is meant to communicate to the victim that the wrongdoer failed to treat them with the proper respect and consideration they deserve. In the cult classic film, *Pulp Fiction*, there is a sequence in the movie where two hitmen, Jules and Vincent, find themselves on the road and in trouble with a dead body in their backseat. The deceased was an associate named Marvin, who was shot accidentally in the head by Vincent. Although Vincent has apologized for the accidental shooting, Jules continues to blame Vincent for the predicament he has put him in. After all, Jules is the one that will eventually have to shoulder much of the responsibility of getting the car off of the road. Vincent takes issue with how Jules continues to blame him, and says, “ Jules, did you ever hear the philosophy that once a man admits that he’s wrong that he is immediately forgiven for all wrongdoings? Have you ever heard that?”¹⁷ Should Jules have forgiven Vincent for his wrongdoing upon his initial apologetic offering? I begin my paper with this quote because it suggests a neglected idea in philosophy that there can be reparative duties not only for wrongdoers, but also for the victims of wrongdoing, assuming that Jules is in fact a victim of Vincent’s fatal mistake. In a recent paper by Cecile Fabre (2024), she argues that victims have a *pro tanto* moral duty to accept their wrongdoer’s apology. This duty involves a doxastic dimension that requires victims to take wrongdoers at their word and believe in the sincerity of their apology. Therefore, the duty to accept apologies implies a doxastic duty to take others at their word when they apologize (p. 366). All things considered, I think this doxastic duty is a fairly intuitive duty

¹⁷ Jules seems to be drawing on an idea found in the passage from Luke 17:4 which states, “...if he sins against you seven times in the day and turns to you seven times, saying, ‘I repent,’ you must forgive him.”

to apply in simple cases of wrongdoing. If you accidentally bump into me causing me to spill my coffee, it seems unreasonable for me to believe you are lying about the apology you have given me unless I have corroborating evidence that speaks to your insincerity. However, as I will explain, in more complex cases of wrongdoing I think we have good reason to doubt the sincerity of our wrongdoer's apology. If I am right about this, one question that naturally arises from these more complex cases of wrongdoing is what our reparative duties are to wrongdoers whom we are justified in not believing? My paper sets out to provide an answer to this question. I will argue that, even when we have evidence to not take our wrongdoer at their word and accept their apology, we can still have a moral duty to *engage* with their apology.

As I will show, this is because apologies can involve a form of moral persuasion. So even if an apology is not initially believed, a wrongdoer ought to be allowed to offer moral reasons to try persuading the victim of his remorse, understanding, and commitment to not reoffend. I will explain that we have a Kantian duty to engage with our interlocutor's moral reasons, and reevaluate our interlocutor's apology in light of those reasons. My argument adds to Fabre's account by providing an explanation of a victim's reparative duties in cases where the victim does not have a duty to believe in the sincerity of her wrongdoer's apology. As I will show, such cases are fairly common given the evidence at the victim's disposal to think that her wrongdoer is not being sincere in his apologetic offering. These cases demonstrate that moral persuasion plays a key role in the ethics of apologies.

A common mistake when thinking about apologies is that they can always be presented in a single declarative statement for the victim to either accept or reject. While this will do the job for minor indiscretions, this will not suffice for wrongs that have fractured the base level of respect and consideration owed to another individual. As I will show, a proper apology—that is an apology that communicates to the victim the wrongdoer's failure to treat them with the proper respect and consideration they deserve—is one that takes effort and nurturing over time. A proper apology will often contain more than the apologetic utterance, but also be preceded and followed by non-verbal ways of communication that include facial expressions, empathy, changes in behavior, and other non-verbal interactions with the victim. These non-verbal interactions provide moral reasons to the

victim that demonstrate the sincerity of the wrongdoer's apology. As I will explain, the provision of moral reasons is more than a means to get someone to change their mind, but it also can constitute a way in which people recognize others as rational and moral agents. When others recognize us as sharers of moral reasons, we have a duty to reciprocate that recognition and engage with those reasons. I argue that in the case of apologies, victims have a duty to engage with the moral reasons provided to them by the wrongdoer and reevaluate the initial apologetic offering in light of those reasons.

My paper will proceed as follows: In section 1, I reconstruct Fabre's argument that we have a *pro tanto* moral duty to take wrongdoers at their word and accept their apology. In section 2, I argue that some wrongdoings, which I call 'complex wrongdoings', provide victims relevant and reliable evidence to doubt the sincerity of their wrongdoer's apology, and thereby justify victims to reject or withhold acceptance of the apology. In section 3, I argue that while victims are justified in rejecting or withholding acceptance of an apology, there remains a moral duty to engage with the wrongdoer's apology when the wrongdoer chooses to provide corroborating evidence for the sincerity of their apology.

1. The Victim's Reparative Duty

Fabre defines an apology as an illocutionary speech act that is vindictive, expressive, and commissive (p. 357). That is, for an apology to count, the wrongdoer must (a) form the correct moral judgment such that he understands the nature of his wrongdoing, (b) must be appropriately remorseful in expressing his apology, and (c) his utterance must commit him to not reoffend.

When a wrongdoer's apology meets these conditions, an apology is only genuinely accepted by the victim if a doxastic requirement is met (pp. 360-61).¹⁸ The doxastic requirement for acceptance calls for the victim to believe that the wrongdoer genuinely understands the nature of his wrongdoing, is appropriately remorseful, and commits to not reoffend. In Fabre's words, the victim "has to believe that he means it" (p. 361). However, mere belief is not sufficient to count towards genuine acceptance.

¹⁸ Fabre also has communicative and attitudinal requirements for acceptance, though for the purposes of my paper I will not explain these here.

Instead, the victim's belief that the wrongdoer is telling the truth must be grounded in the wrongdoer's apology. That is, what counts as relevant evidence for the victim's belief that the wrongdoer understands the nature of his wrongdoing, is appropriately remorseful, and commits to not reoffend, will have to be informed—at least in part—by the wrongdoer's self-avowal.

A wrongdoer is typically thought to have a reparative duty to their victim in light of failing to treat their victim with the respect and consideration they deserve. As I will use the term, a 'reparative duty' will refer to a duty to repair the break in the proper treatment of someone worthy of basic respect and consideration. The reparative work of an apology is to signal to the victim that they are in fact deserving of the wrongdoer's concern and respect, and that the wrongdoer failed to treat them in this way. A wrongdoer compounds their initial wrongdoing when they fail to apologize and make proper amends to their victim (p. 359).

At first blush, it might seem strange to think the victim can *also* have a reparative duty to the wrongdoer, given they are the ones who have been wronged. However, with the above definitions of apology and acceptance in mind, Fabre argues that victims have a *pro tanto* moral duty to their wrongdoer to accept their apology. This is because there are constraints to the hard treatment victims can justifiably put wrongdoers through. Such constraints are instantiated by a Kantian imperative to give the respect persons deserve as rational and moral agents (pp. 363-64). This imperative explains the duty of the victim to take the wrongdoer at their word and accept their apology. Respecting a person as a rational agent involves treating them as fellow reasoners, capable of engaging in both the provision of reasons as well as the ability to respond to them. Respecting a person as a moral agent involves recognizing a person's capacity to judge right from wrong and to expect from them a willingness to take responsibility in light of those judgments. An implication of this imperative is that "we may deliberately subject someone to hard treatment only if and on the grounds that they have acted in such a way as to warrant it; or, under conditions of uncertainty, only if and on the grounds that we have sufficient evidence to that effect" (p. 364). This is because respecting persons as rational and moral agents calls on us to extend to them a presumption of innocence, absent relevant and reliable evidence of their guilt. But just as there is a duty to withhold hard treatment of others absent proper evidence of

their guilt, then *a fortiori* the same duty applies when others provide proper evidence of their innocence (p. 365). If a victim continues to subject a wrongdoer to hard treatment after they have provided relevant and reliable evidence of their understanding, remorse, and commitment not to reoffend, then the victim fails to treat the wrongdoer with the proper respect they deserve as a rational and moral agent.

According to Fabre, this imperative also applies to doxastic hard treatment, where we are constrained not only in our treatment of others, but also in our beliefs of others. We are not morally permitted, for example, to negatively stereotype others based on unfounded assumptions about their race or gender. Similarly, it would be wrong to maintain a morally unsavory belief of someone for wrongdoing after they have provided us with relevant and reliable evidence of their sincere remorse, understanding, and commitment to not reoffend. The duty to withhold doxastic hard treatment of others, according to Fabre, implies the duty to believe others when they apologize, absent relevant and reliable evidence that give us reason to doubt them. In Fabre's words, "Unless we have evidence that he is not to be trusted, we are under a duty to recognize him as having testimonial authority in respect of the facts that concern him" (p. 365). Therefore, if we assume the wrongdoer's apologetic utterance meets the vindictive, expressive, and commissive requirements explained previously, then the victim has a duty to take the wrongdoer at their word and accept their apology, absent evidence to justify not believing in the wrongdoer's remorse, understanding, and commitment to not reoffend.

2. Complex wrongdoings

What should we make of the duty to take the wrongdoer at their word and believe in their remorse, understanding, and commitment to not reoffend? In many cases of wrongdoings, I think the duty can be applied without much difficulty. Take a case that Fabre gives us, where Walter is driving at a recklessly high speed and accidentally crashes into Violet (p. 358). Walter's apology to Violet includes a genuine acknowledgment of his remorse, understanding, and commitment to not reoffend. Unless Violet has evidence to not take Walter at his word, it would seem that Violet should believe Walter and

accept his apology. After all, there doesn't seem to be any apparent reason available to Violet to think that Walter's apology is insincere. However, it strikes me that not all wrongdoings are as straightforward as this one, and that other wrongdoings will in fact provide evidence to doubt the wrongdoer's remorse, understanding, and/or commitment to not reoffend. My aim in this section is to explain that wrongdoings that are *intentional*, *cause severe harm*, and/or are *repeat offenses*, can provide evidence to the victim to doubt the sincerity of the wrongdoer's apology, and therefore justify the victim to withhold acceptance of the apology. I refer to these wrongdoings as 'complex wrongdoings.' Even if the wrongdoer is indeed sincere in his apology, the evidence for this will not always be clearly available to the victim in cases of complex wrongdoings. To be clear, my aim is not to dispute the pro tanto moral duty Fabre claims we have to accept apologies. Both Fabre and I agree that when a victim has relevant and reliable evidence to doubt the sincerity of the wrongdoer's apology, the victim is then under no pro tanto duty to accept that apology. The upshot of this section is to help delineate between wrongs that provide that evidence from those that do not.

Intentions:

One way in which wrongdoing can be 'complex' is when the wrongdoer *intentionally* does something to the victim *with the understanding that it is wrong*. That is, I think a wrongdoer's intentions can create a reasonable doubt for the victim when asked to believe outright the sincerity of the wrongdoer's remorse. We can define remorse as expressing and feeling regret, sorrow, and repentance for what we could have done differently but chose not to do, or were culpably negligent for not doing. As explained in the last section, Fabre argues that the victim has a duty to accept the wrongdoer's apology by believing that the wrongdoer is sincerely remorseful. I argue that a wrongdoer's intentions behind a wrongdoing can nullify that duty. To see how, consider the following case:

Sabotage: Law firm co-worker's Luke and Kiera learn their law firm has an opening for a junior partner position. As the two longest tenured associates, both are at the top of the list for

consideration. The two have been friends during their tenure, but Luke decides to secretly sabotage Kiera's chances of getting the promotion to better position himself. During the months leading up to the firm's decision, Luke spreads false rumors about Kiera to other associates, some of which are relayed to the senior partners. Luke also tampers with Kiera's work, causing setbacks in some of her client's cases which ends up reflecting poorly on her professional reputation. Eventually Kiera finds out about Luke's plans to sabotage her chances at junior partner, and she confronts him. Luke admits his wrongdoing, and apologizes to Kiera.

Let us suppose that Luke is telling the truth and is sincerely remorseful, understanding of his wrongdoing, and committed to not reoffend. In other words, Luke does in fact apologize according to Fabre's definition. The question is, however, whether *Kiera* has evidence at her disposal to doubt the sincerity of Luke's apology? I argue that she does. More specifically, the intentions behind Luke's wrongdoing give Kiera a reason to justifiably doubt the sincerity of his remorse. This is because Kiera has to reconcile the unremorseful person Luke was at t_1 with the person that now apologizes at t_2 . Because he intentionally did something to wrong Kiera, Luke has betrayed her trust in a way that Walter does not with Violet when he unintentionally runs into her. In regards to Luke at t_1 , if we assume he understood his actions and acted intentionally, it would seem he was not remorseful, but rather unrepentant and remorseless. Moreover, Luke has given evidence to how he can be intentionally deceptive and manipulative in order to get what he wants at the expense of Kiera's interests. The need for reconciliation of character is not obvious in the case of Walter, in large part because Walter did not intentionally harm Violet. Walter's recklessness, though blameworthy, does not seem to be a character defect that is antagonistic to feelings of remorse. That is, it does not seem like Walter at t_1 gives evidence to doubt his remorse upon his apologizing at t_2 —at least not in the same way Luke does. It seems like part of what is wrong with Luke's behavior stems from him knowingly harming Kiera and doing it anyway. In order to be genuinely remorseful, Luke will need to give Kiera's interests more weight than he had earlier. Intentionally deceiving Keira and causing her harm indicates not only a

striking lack of compassion, but also character attributes that are antagonistic to feelings of remorse, such as a maliciousness and a general apathy for her wellbeing. I don't think these character attributes can always be adequately reconciled through a straightforward apology. If I am right about this, I think Kiera is justified in withholding acceptance due to a legitimate skepticism about Luke's remorse given the evidence she has of his character at t1. To take it a step further, we might even think it naive and reckless of Kiera to immediately believe in Luke's remorse despite his intentional mistreatment of her. Regardless, I do not think it should be morally required of her to believe in his remorse upon his apology.

Repeat wrongdoing:

As with intentions, I think cases of repeat wrongdoings provide evidence to the victim to justifiably doubt the sincerity of the wrongdoer's apology. Specifically, I argue that repeat wrongdoings provide evidence to doubt the wrongdoer's commitment to not reoffend. Take the following case:

Crash: Race car driver, Max, has a reputation amongst his peers as a reckless driver. Although Max never intentionally causes accidents on the race track, he is prone to attempting reckless maneuvers that other drivers would never attempt out of concern for the safety of others. During a race, he tries a characteristically reckless maneuver in an effort to overtake Stewart, causing Stewart to crash into the barrier. This is the second time Max has caused Stewart to crash in the last four races due to a reckless maneuver. After the race, Max apologizes to Stewart, and tells him that it won't happen again.

The evidence to doubt the sincerity of Max's apology can be found in his history of repeating the very same offense he is apologizing for. Unlike first time offenders, Max has provided evidence for patterned behavior in regards to reckless driving which gives reason to doubt the truthfulness of his commitment to not reoffend. This is not to say that Max will in fact commit the wrongdoing again, nor is it to say that he is in fact insincere in his apology—let us assume that he in fact is telling the truth

when he commits to not reoffend. Rather my claim is that the evidence at the victim's disposal is such that a justified skepticism of the wrongdoer's commitment to not reoffend is warranted.

Fabre agrees that repeat wrongdoings can epistemically justify the victim to forgo accepting the wrongdoer's apology (p. 372). However, as Fabre points out, one might object that Stuart wrongs Max by not accepting his apology, if in fact Max is sincere in his commitment to not reoffend. This type of wrong is what Derek Parfit (2011) refers to as a 'fact-relative wrong', and can be distinguished from 'evidence-relative wrongs' (*On What Matters*, pp. 150-162).¹⁹ To explain the difference between the two, suppose you are at a dinner party in which you have relevant and reliable evidence that your host intends to poison the food. However, despite the strong evidence at your disposal, your host in fact does *not* intend to poison the food. In this case, assuming you were unaware of the facts, you would be justified in an evidence-relative sense in accusing the host of intending to poison the food. However, you would not be justified in a fact-relative sense in accusing your host, and by doing so would be committing a fact-relative wrong, since your host does not actually intend to poison the food. In the case of *Crash*, if Stewart does not believe in Max's commitment to not reoffend, and Stewart is in fact making a sincere commitment to not reoffend, then Stewart seems to commit a fact-relative wrong when he fails to accept Max's apology. Now if there is a fact-relative wrong, then it's another question as to whether the harm is directed at the wrongdoer, or if the harm is undirected. If the harm is undirected, perhaps the victim is not culpable for the fact-relative wrong. But, as Fabre explains, even if the harm is directed, this wrong must be explained using a different account than the one she provides (pp. 372-73). As explained in section 1, Fabre argues that the victim wrongs the wrongdoer when she fails to accept the wrongdoer's apology absent evidence to doubt the sincerity of the apology. In the case of Stuart, the evidence the victim has at his disposal is such that he is justified in not believing Stewart, and therefore does not wrong him according to Fabre's doxastic requirement. For this reason, I will not pursue further the possibility of the victim committing a fact relative-wrong against the wrongdoer, but rather conclude that repeat wrongdoings can provide sufficient evidence for the victim to justifiably withhold believing the wrongdoer's apology.

¹⁹ Fabre uses Parfit to make this point in her paper on p. 372.

Severe harm:

According to Fabre, understanding the harm one causes requires the wrongdoer to acknowledge both the nature of the harm as well as their role in the harm caused to the victim (p. 358). As I understand the former part of the requirement, part of what it means to understand the nature of the harm is to understand the degree to which the victim has been harmed. For example, if I culpably crash into you with my car and cause you to break your arm, I fail to understand my wrongdoing if I understand the harm caused to you as equivalent to the harm you might endure in a simple fender bender. In order for my apology to count, I must acknowledge the degree of harm inflicted upon you, as well as my role in inflicting said harm. I claim that in cases of severe harm, the victim can justifiably doubt the wrongdoer's understanding of the degree of harm they have caused, as well as their understanding of their role in said harm. Take the following example:

Belay: Joe is belaying Mark on a fifty foot wall at an indoor gym. The two are both climbing enthusiasts, and are familiar with each other's climbing styles. Mark is about thirty feet up the wall when Joe starts a conversation with another climber who has just entered the gym. Distracted by his conversation, Joe does not realize that he has given the rope far too much slack. At thirty-five feet, Mark tries an advanced move on the rock and falls. Because Joe is distracted and has given the rope too much slack, he fails to make a routine catch which causes Mark to drop to the ground and break his back, most likely preventing Mark from ever climbing again.

I claim that the degree of harm caused in the above example can be harder to capture than in cases of minor harm, such as the harm experienced in a mere fender bender. For Joe to sufficiently understand the harm he has caused, he must in some sense understand what rock climbing means to Mark. More specifically, Joe must understand what it means for Mark to never be able to rock climb again. I take *Belay* to be illustrative of how the degree of harm can generally be more difficult to grasp

in cases of severe harm, as we often don't know what it would be like for another when said harm is inflicted upon them. On the other hand, when it comes to causing less severe harm, like in the case of a fender bender, I think our general familiarity with such scenarios—even if we have never been a victim of a fender bender— can allow us to get sufficiently close to understanding what it would be like to experience said harm. With severe harm it's not always so simple, as our guess of what it would be like can be less accurate due to our unfamiliarity with scenarios involving more severe harms. Moreover, I think it can be difficult for the wrongdoer to come to terms with how *they* could be responsible for inflicting severe harm—intentional or otherwise—to another person. This is because causing severe harm generally does not fit a familiar script of responsibility. With severe harm, it can take time and reflection for the wrongdoer to come to terms with their role in the harm they have caused. This is perhaps part of the reason for why an immediate apology after a severe harm seems superficial. It is simply hard to believe that a wrongdoer that causes significant harm can fully appreciate their role in the harm they have caused the victim in such a short period of time. This temporal requirement for understanding one's role in causing severe harm is perhaps linked to our need for narrative and storytelling to make sense of the world and our roles within it (Breithaupt, 2025). While already familiar narratives can usually be weaved into the stories we tell ourselves without much difficulty, new narrative frameworks, particularly those that are negatively valenced, can often take more effort and time to integrate into our own personal story (Hartog et al., 2020). I think integrating new and negatively valenced experiences can be especially difficult in cases where one is blameworthy for causing another severe harm. While minor wrongs such as accidentally causing a fender bender are fairly compatible with the narrative that people are fallible and are prone to sometimes inconveniencing others, severe harms such as the one Joe is responsible for do not easily map onto the stories we tell ourselves about ourselves. Even though it was accidental, Joe will have to find a way to reconcile his role in preventing Mark from ever climbing again with the person he saw himself as prior to the accident. That causing severe harms are formative in a way that causing minor harms are not is a fairly intuitive idea. After all, I think we would have good reason to doubt Joe's basic understanding of the harm he had caused if he emerged from *Belay* emotionally unscathed and psychologically unaffected. Such

experiences should, and typically do, leave an impression on the wrongdoer who takes the time to genuinely process his role in significantly impacting another person's life. Therefore, if we assume that the degree of harm caused to Mark lies outside the typical range of harm for which Joe has previously found himself responsible for, then we should also think it is likely to take Joe additional time to reconcile the blame for Mark's injury with his own personal story.

My point is not that it is impossible to understand severe harm. Rather, I only claim that understanding severe harm can generally be more difficult and time consuming than with less severe cases. Given this difficulty, I claim the victim can be justified in having a reasonable doubt about the wrongdoer's understanding of the harm caused, and thus can be justified in withholding acceptance of the wrongdoer's apology.

3. Apology and persuasion

In the last section, I argued that wrongdoings that are intentional, cause severe harm, and/or are repeat offenses, can provide evidence to the victim to doubt the sincerity of the wrongdoer's apology, and therefore justify the victim to withhold acceptance of the apology. However, I think there still can be a reparative duty on the part of the victim in these cases. I think a common mistake that is made when thinking about apologies is that the victim can always be sure of whether to accept or reject an apology on its face. I think sometimes wrongdoings can be complex, and will require a wrongdoer to make their case to the victim as to their sincerity regarding their remorse, understanding, and commitment to not reoffend. Moreover, a convincing case by a wrongdoer can not always be made in a single declarative statement that can be accepted or rejected, but rather will require taking place during a fluid interaction between wrongdoer and victim over time, where a wrongdoer provides evidence of his sincere apology that can then be deliberated on by the victim.

To see how an apology might take place over time, let us acknowledge at least two possibilities on the part of the wrongdoer after his or her apology has not been accepted: (1) the wrongdoer sees that his apology has not been accepted and moves on OR, (2) the wrongdoer is either asked by the

victim to follow-up and provide further evidence to corroborate his apology and the wrongdoer agrees, or the wrongdoer decides (unsolicited) to follow-up with the victim to see if they will accept the apology given the corroborating evidence. In this section I argue that in case (2), victims have a moral duty to engage with the corroborating evidence provided by the wrongdoer, and reevaluate the wrongdoer's apology in light of said evidence. The upshot of my argument is that even when we are not under a moral duty to accept a wrongdoer's apology, we can still have a moral duty to *engage* with their apology.

3.1 corroborating evidence post-apology

How might the wrongdoers in the cases explained in the last section provide evidence to their victims to help corroborate their apologies? Let us start with *Sabotage*. One main difference from the case Fabre gave us of Walter seems to be that Luke's remorse will take more time to prove, such that proving his remorse will not be captured in a single apology but rather over a multitude of actions post-apology. Of course Luke should apologize first, despite lacking evidence for his sincere remorse. He owes Kiera an apology, but this is only the first step. Following the apology he must demonstrate a rehabilitation of his character from when he committed the wrongdoing. This will include a transformation of character expressed not just through words, but through rehabilitative and restitutive actions post-apology. For example, Luke needs to show through his action that he no longer is willing to deceive and lie for his own benefit. This will involve sincere and compassionate interactions with Kiera and colleagues over time. Moreover, he will have to help with correcting the lies and the damage he has caused. For example, he will have to clear things up with the other associates and the junior partners and clarify that the rumors he spread were not true, but rather fabricated by him. Moreover, in rehabilitating his character post-apology, Luke must be a willing participant in reforming his character. That is, he must go about his rehabilitation with the proper attitude. For example, it seems problematic for Luke to express reluctance during his rehabilitation. A willing attitude can be expressed through his general demeanor, facial expressions, and body language.

In the case of *Crash*, Max would have to demonstrate a clean record of driving over time, particularly in moments where he was characteristically reckless. Like with *Sabotage*, because Stewart has reason to doubt the sincerity of Max when he apologizes, what is needed are actions that speak to a change in his character, and particularly to his commitment to not reoffend.

In *Belay*, in order for Joe to persuade Mark that he is sincerely understanding of his wrongdoing, he should make a concerted effort to not repeat his mistake with other climbers. Because Joe's mistake was caused by his lack of concentration, he should demonstrate a heightened focus when belaying future climbing partners. For example, if Joe continues to be prone to distractions while belaying his partners—even if no further accidents occur—this not only signals a higher probability of reoffending, but also a failure to grasp the significance of the harm caused to Mark.

Moreover, Joe should show Mark that he is making a concerted effort to understand the harm he has caused and reflect on the impact the accident will have on Mark moving forward. For example, this may include having conversations with Mark to better understand his experience regarding both the incident and his recovery. As explained in section 2, one of the reasons why Joe may fail to fully understand the harm caused to Mark is because of the general difficulty a wrongdoer faces in understanding how severe harms such as *Belay* may impact the victim. Therefore, to persuade Mark that he fully understands the harm he has caused, Joe will need to make a concerted effort to *empathize* with his experience. As Olivia Bailey (2020) explains, empathy serves as a unique source for what she calls “humane understanding”, which she defines as the apprehension of the first-hand emotional intelligibility of the other (pp. 8-9). For example, to apprehend Mark's emotional state following the accident first-hand, Joe would not only need to imagine how he would feel in Mark's position, but rather he would need to imagine occupying his position through Mark's emotional lens (e.g. the emotions that accompany the thought of never climbing again. Only by taking a first-hand perspective of Mark's emotional experience can Mark's emotions become intelligible to Joe. This is because what is needed for intelligibility is more than strictly an intellectual appreciation of the evaluative properties of emotions. As Bailey puts it, “Judgments that things have certain evaluative properties can be cold” (p. 8). In order for Mark's emotional experience to be intelligible to him, Joe must not simply approve of

the evaluative properties of Mark's emotional experience, but must engage directly with his emotional experience.

Why is "humane understanding" valuable? Bailey understands its value in the effects of its absence. When we are not humanely understood, we can feel as though our emotional experiences are alien, and only intelligible to ourselves. Bailey describes this feeling as "a particular form of anxiety" that cannot be treated by the mere approval of our emotional outlook. Simply approving of our emotional experience can, while perhaps less isolating than rejecting our emotional testimony altogether, leave us alienated in a position of authority (p. 11). I think that a wrongdoer's effort to genuinely engage in humane understanding is especially important for victims when they are considering a wrongdoer's apology. This is in part because the victim and the wrongdoer stand in a special relation to one another as the primary actors in a shared experience. It seems especially alienating if the only other person whom you stand in special relation to regarding a shared experience fails to empathize with your emotional state regarding that experience. The alienation that comes from not receiving humane understanding in a shared experience seems especially potent when considering the experience of being wronged. This is because when a wrongdoer does not even attempt to show humane understanding with regard to his victim's experience, he signals a refusal to share in the first-hand emotional experience of the harm he was culpably responsible for causing. A refusal or obstinacy to empathize with the victim and engage with her first-hand experience suggests to the victim that, although the wrongdoer may approve of her emotional state by trusting that she is feeling the appropriate emotions, he is unwilling to do the moral work necessary to fully understand the harm he has caused her. Therefore, by failing to empathize with his victim, the wrongdoer not only leaves his victim emotionally alienated in an emotional experience he is culpably responsible for causing, but also conveys that he does not regard her as a moral agent worthy of respect and proper consideration.

3.2 Moral persuasion

In all of these cases, I take the provision of corroborating evidence to constitute instances of ‘moral persuasion.’ I define moral persuasion as the provision of reasons aimed to change the target’s mind on some moral matter. The moral matter in question when wrongdoers attempt to persuade the victim to accept their apology is if the wrongdoer is sincerely remorseful, understanding, and committed to not reoffend. An important element of my definition of moral persuasion is that the provision of reasons is not limited to facts and the straightforward exchange of information. Instead, I refer to moral persuasion as the conveyance of any evidence that is relevant to the moral deliberation of the victim in deciding whether the wrongdoer sincerely understands his wrongdoing, is remorseful, and is committed to not reoffend. The relevant evidence for the victim’s moral deliberation can be communicated in a variety of different ways, including, but not limited to, the wrongdoer’s testimony, actions, expressions of emotion, empathy, and general non-verbal interactions with the victim. I argue that if the wrongdoer is willing to provide corroborating evidence and in fact does, then the victim has a moral obligation to open themselves to the wrongdoer’s attempt at moral persuasion. To open oneself to moral persuasion in these cases is to be understood as sincerely engaging with the corroborating evidence and reevaluating the apology in light of said evidence.

3.3 two qualifications

Before explaining the duty to open oneself to moral persuasion, two qualifications are in order. (1) So far I’ve used the term ‘corroborating evidence’, while the definition of moral persuasion has to do with the provision of ‘reasons’. Therefore, one could object that actions, demeanor, and other non-verbal signals, while counting as relevant corroborating evidence, might not be thought of as communicating *reasons* that help to explain why the target should adopt a particular position on a moral matter. However, moral persuasion can be communicated through other means than just the provision of verbalized reasons when non-verbalized evidence is contextualized (Rini, p. 17-18). This is particularly true when it comes to apologies, since we communicate through non-verbal means all the time in the context of post-apology work such as restitution and rehabilitation, and thus we are

familiar with how such evidence translates into reasons. For example, when I try making amends through action after an apology, it seems intuitive that my actions corroborate my initial apologetic offering, and are interpreted as providing supporting reasons for accepting my apology. Much of the relevant and familiar evidence that speaks to remorse, understanding, and likelihood of not reoffending, will come through non-verbal expression post-apology. When this evidence is presented in the context of what follows from an apology, it can be assumed to be given and received as reasons that speak to remorse, understanding, and not reoffending. Therefore, while the above examples provide evidence rather than verbalized reasons, I think it is reasonable to assume this evidence can be interpreted as giving reasons to the victim to accept the wrongdoer's remorse, understanding, and commitment to not reoffend. Henceforth, when I use the term 'evidence', I will assume that such evidence translates into reasons that can be recognized as such, and be deliberated on by the victim.

(2) While the victim has some discretion as to what evidence is required for them to be convinced of the wrongdoer's remorse, understanding, and commitment to not reoffend, the victim cannot ask the wrongdoer to debase themselves or compromise their dignity. Although they can push back and question the corroborating evidence the wrongdoer provides, there are limits to what the victim can expect from the wrongdoer post-apology. For example, the victim should not ask the wrongdoer to repeat reparative actions with the aim of embarrassing or shaming them. The duration of the persuasive attempt also cannot be extended past what is reasonable. That is, the victim at some point must decide on whether to accept or reject the wrongdoer's persuasive attempt. Neither should the reasons behind requesting more evidence be motivated by vitriol or animosity, but rather it should be limited to obtaining a sufficient amount of evidence needed to make an informed decision on whether to accept the wrongdoer's apology. In short, any request the victim makes on the wrongdoer to provide corroborating evidence must be within reason and be motivated by looking to deliberate on the sincerity of the wrongdoer's apology.

3.4 The duty to engage with moral persuasion

Unlike the use of other methods of influence like manipulation and coercion, when I try to morally persuade you, it matters to me that you change your mind for the reasons I give you. Therefore, to be open to my attempt at moral persuasion will require you to sincerely engage with my reasons and not simply change your behavior in order to, say, appease me. Similarly, it would be equally inadequate if you reject my attempt to persuade you without engaging with the supporting reasons I have provided. On the other hand, it is consistent with my aims at morally persuading you if, after having sincerely engaged with the reasons I have offered, you decide to reject my moral view because you are not convinced. As Regina Rini (2018) explains, I expect that you engage with the reasons I have offered you because when I attempt to morally persuade you I treat you as a sharer of reasons, and not just an object to be manipulated and controlled. That is, I treat you as an agent capable of recognizing and responding to reasons that do not merely advance self-interested aims, but are meant to apply to all agents no matter the scenario (p. 5). As Rini points out, the type of respect that moral persuasion conveys is what Stephen Darwall (2009) calls ‘recognition respect.’ According to Darwall, “Recognition respect concerns, not how something is to be evaluated or appraised, but how our relations to it are to be regulated or governed. Broadly speaking, we respect something in the recognition sense when we give it standing (authority) in our relations to it” (p. 123). When you decide not to engage with my reasons, you fail to reciprocate the recognition respect that I extended to you as an authoritative sharer of moral reasons. That is, when you decline to engage with my reasons, you communicate to me that I do not have standing as a reason giver, and therefore I do not hold equal status in relation to you as a member of the moral community. Insofar as we have a moral duty to reciprocate recognition respect to fellow agents, we also have a duty to open ourselves to an agent’s attempt at moral persuasion.

If I am right that efforts to persuade the victim to accept the wrongdoer’s apology are instances of moral persuasion, then we have a moral duty to engage with the reasons the wrongdoer provides through corroborating evidence post-apology. I want to emphasize that opening oneself to moral persuasion does not necessitate accepting the moral view of the other. As a form of moral persuasion, the victim does not have an obligation to accept the apology in light of the evidence provided, but

rather is only obligated to engage with the evidence and reevaluate the apology in light of said evidence. For example, in the case of *Sabotage*, Kiera is only obligated to engage with Luke's post-apology efforts, but she has no obligation to accept his apology in light of these efforts. When it comes to moral persuasion, the duty only necessitates the engagement with the others' reasons, rather than the acceptance of their argument (Rini, p. 6). That is, it is consistent for the victim to open themselves to the wrongdoer's reasons without accepting their moral view. As previously explained, the reason for why a wrongdoer might provide corroborating evidence to the victim is because the victim has reason to doubt the sincerity of the wrongdoer's apologetic offering. The reasons for doubt are provided in the wrongdoer's intentions, history of reoffending, and/or the severity of harm inflicted upon the victim. Therefore, in many cases of wrongdoing the victim will not have an obligation to accept the wrongdoer's apology, but will still have a duty to engage with it.

Conclusion

An important point that Fabre's argument sheds light on is that reparative efforts often will require more than just an effort from the wrongdoer, but also the victim. Like Fabre, I too believe that reparative responsibilities should be placed on both wrongdoer and victim. However, sometimes the wrongdoer's offense can affect the victim's trust in the wrongdoer, requiring the wrongdoer to regain the trust of the victim, specifically following a complex wrongdoing. While in these cases I agree with Fabre that there is no duty on the part of the victim to believe the wrongdoer and accept their apology, I have argued there is still a duty for the victim to engage with the apology as a form of moral persuasion. What makes this significant is that my argument shows why apologies can still deserve consideration in cases where victims are justified in doubting the sincerity of their wrongdoers. The argument I have defended builds on Fabre's acknowledgement of the reparative responsibilities of both parties by demonstrating the key role moral persuasion has in the ethics of apologies. This role is made clear when we acknowledge that apologies often involve reparative efforts over time, and are often not

simply a declarative statement that can be easily accepted or rejected. Because these efforts consist in the provision of moral reasons, the victim has a duty to engage with them.

References

- Bailey, O. (2020). Empathy and the Value of Humane Understanding. *Philosophy and Phenomenological Research*, 00:1–16. <https://doi.org/10.1111/phpr.12744>
- Breithaupt, F. A. (2025). *The Narrative Brain: The Stories Our Neurons Tell* (1st ed.). Yale University Press.
- Fabre, C. (2024). The Duty to Accept Apologies. *Journal of Moral Philosophy*, 21(3–4), 355–378. <https://doi.org/10.1163/17455243-20234057>
- Hartog, I., Scherer-Rath, M., Kruijzinga, R., Netjes, J., Henriques, J., Nieuwkerk, P., Sprangers, M., & van Laarhoven, H. (2020). Narrative meaning making and integration: Toward a better understanding of the way falling ill influences quality of life. *Journal of Health Psychology*, 25(6), 738–754. <https://doi.org/10.1177/1359105317731823>
- Parfit, D. (2011). *On What Matters* (Vol. 1). Oxford University Press. <https://doi.org/10.1093/acprof:osobl/9780199572809.001.0001>
- Tarantino, Q. (Director). (1994). *Pulp Fiction* [Film]. Band Apart; Jersey Films; Miramax Films.
- Rini, R. (2018). Abortion, Ultrasound, and Moral Persuasion. *Philosophers' Imprint*, 18(6): 1–20. <http://hdl.handle.net/2027/spo.3521354.0018.006>
- Darwall, S. (2009). *The Second-Person Standpoint: Morality, Respect, and Accountability*. Harvard University Press.