

Paternalism and Evidence of Incapacity: Taking Reasons Seriously

김수진 (경희대학교)

Abstract

One of the most salient objections against paternalism is that it is motivated by a negative judgment about other people's capacity to advance their own goals and interests. Such a negative judgment, according to this objection, is morally wrong because it denies others the status of moral equals who can rationally set and pursue their own conception of the good. Despite the popularity of this objection, I argue that it misfires because rendering a negative judgment about others' capacities does not deny them equal moral status when there is sufficient *evidence* for that judgment. However, I argue that in order to determine if there is sufficient evidence of the agent's incapacity, we must understand the agent's reasons for action through a rational discursive exchange with them, instead of taking perceived facts about their behavior as sufficient grounds for a negative (or positive) judgment of their capacities. In fact, I argue that if the outcome of the discursive exchange supports a negative judgment of the agent's capacities, there is a pro tanto reason in favor of paternalistic interference.

Introduction

One of the most salient objections against paternalistic state action is that it expresses contempt and disrespect for citizens' capacities to advance their own interests. The claim is that imposing paternalistic policies on citizens involves treating citizens as if they cannot rationally promote their own conception of the good, which amounts to denying them equal moral status (Quong 2010, p.101; Shiffrin 2000, p. 220; Dworkin 1988; Tsai 2014; Cornell 2015).¹

Yet, others have challenged this *Equal Respect objection* to paternalism on the grounds that it merely indulges people's wishful thinking about how they might like to be regarded by others. This view, which is grounded in the epistemological theory called *evidentialism*, says that if facts about the agent's behavior seem to show that the agent tends to deliberate irrationally, misconstrue the relevant evidence, or be weak-willed, then the right thing to believe is that the

¹ Relational egalitarians also sometimes object to state paternalism on similar grounds. See Anderson 1999, p. 301; Schemmel 2011, p. 124.

agent is incapable of effectively advancing her own interests.² These conflicting views raise the following two questions: does evidence of a person's incapacity cancel out the expressive wrongs of paternalistic behavior, and what constitutes "evidence" of a person's incapacity?

I argue that although sufficient evidence of a person's incapacity cancels out the expressive wrongs of paternalistic behavior, "sufficient evidence" of a person's incapacity should not be conceptualized as self-explanatory, observational facts about the agent's behavior, as some assume.³ To assume that the evidence concerning the agent's capacity to promote her own interests is similar to the observational evidence concerning the agent's height is tantamount to ignoring the agent's reasons for action. If we presume that the agent is a moral equal who can make sound judgments about her own reasons, then we must take the agent's reasons seriously as considerations intended to rationally justify her action. But taking the agent's reasons seriously in this sense requires us to try to understand why the agent was led to see her reasons as *good* grounds for her action, instead of concluding that she is irrational or mistaken based on our unilateral judgment of her behavior.

But the process of understanding another agent's reasons, I show, entails engaging in a discursive exchange with the agent herself, since we cannot avoid relying on our own standard of judgment to make sense of the agent's reasons (at least initially) and therefore must argue out with the agent whose standard of judgment is better justified.⁴ A merely hypothetical process of idealization conducted within one's own mind cannot determine whose standard is better

² Enoch 2016; Fox 2019, pp. 325-326. For an overview of evidentialism, see Conee and Feldman 2004.

³ See Enoch 2016, pp. 25-26, where he says that "there is no difference" between evidence of a person's comparative intelligence and evidence of a person's comparative height.

⁴ I will explain in more detail about what I refer to as the "standard of judgment" in the main body of the paper. Roughly, I mean a standard for judging whether reasons for a particular action or belief are sound or unsound, good or bad, right or wrong.

justified since thinking about what a rational person would choose to do in hypothetical, ideal circumstances is compatible with various conflicting judgments.

The claim I seek to establish is not that the outcome of discursive exchange in non-ideal circumstances is always correct or decisive. Instead, the core claim of this article is that in order to reach a warranted judgment about another agent's capacity to promote her own interests, we must first understand why that agent was led to see her reasons as sufficient grounds for her action through a discursive exchange with her.

This article is structured as follows. First, I sketch out the Equal Respect objection against paternalism, and show why evidentialists are right to argue that the Equal Respect objection misfires when there is sufficient evidence of the agent's incapacity. Second, I argue that although evidentialists are correct that the evidence concerning *p* should determine what we are justified in believing about *p*, their conceptualization of the "evidence" concerning another person's (in)capacity does not take seriously enough that person's reasons for action. Specifically, I show that a proper conceptualization of "evidence" must be sensitive to the distinction between (1) *interpretative* questions which require understanding the agent's reasons for action, and (2) *non-interpretative* questions which do not. I claim that because the evidentialist thesis collapses this distinction, it raises both ethical worries about condoning unfair and discriminatory judgments about other people's abilities, as well as epistemic worries about legitimating inaccurate judgments about others' abilities.

Third, I address the question that the arguments of the second section raise: namely, what is required in order to understand other people's reasons for action, and what implications does this have for our judgments about others' capacities to advance their own interests? I show that in order to properly understand an agent's reasons, we must engage the agent in a discursive

exchange aimed at reaching a consensus as to why she thinks her reasons are sufficient grounds (or justification) for her action. On the basis of this argument, I claim that a warranted judgment of the agent's capacities must be determined by the outcome of a discursive exchange with the agent herself.

Finally, I build on the preceding argument to conclude that if the outcome of a discursive exchange with the agent under suitable conditions shows that she is incapable of advancing her own interests on terms that she cannot reject without contradicting herself, then paternalistically interfering with her does not deny her equal moral status, and that this establishes a *pro tanto* (but not an all-things-considered) case for interference.

1. The Equal Respect Objection to Paternalism

Although there is a lively debate about the correct definition of "paternalism," it is not necessary to rehearse that entire debate here since the specific feature of paternalism that this article addresses has to do with what many have alleged to be the main wrong-making feature of paternalism: a negative *judgment* or *belief* about other people's capacity to advance their own interests. According to proponents of this view, what makes paternalism wrong is not the act of restricting another's liberty without her consent, but instead, the insulting *judgment* that the other is too obtuse or weak-willed to do what's best for herself.

Jonathan Quong, for instance, claims that paternalism is inconsistent with the liberal principle of respecting others as moral equals because a moral equal, by definition, must be capable of rationally advancing their conception of the good on their own (Quong 2010, pp. 100-106). Specifically, Quong claims that since paternalistic action is always based on a "*negative judgment* about the paternalizee's capacity to effectively advance his or her ends (emphasis

mine),” treating others paternalistically expresses a denial of their equal moral status (Quong 2010, p. 83). Since the core of this objection is that the belief or judgment that motivates paternalistic behavior is disrespectful to others’ status as equals, I will call this the *Equal Respect objection* to paternalism.

But according to some, the Equal Respect objection is misplaced because it is merely based on a wishful thinking that is incongruent with the facts about people’s capacities. If all the relevant evidence indicates that someone is not very good at advancing her own interests due to her inability to comprehend the facts, lack of reasoning power, or weakness of will, then it seems like there could be nothing wrong – not even *prima facie* wrong – with forming a negative judgment about her capacities. David Enoch, for instance, argues that “The judgments about others’ likely behaviour are governed by epistemic norms, and even if those are somewhat morally sensitive, still when the evidence (sufficiently) supports the judgment that someone is likely to deliberate irrationally, or to be weak-willed, or some such, that is precisely the thing to believe” (Enoch 2016, pp. 39-40). Call this the *Evidentialist thesis*, since this view is based on the epistemological view called “evidentialism” which says that epistemically justified beliefs are determined solely by the relevant factual evidence (Conee and Feldman 2004; Enoch 2016, pp. 25-26).

Quong anticipates something similar to the Evidentialist thesis, but fails to give a satisfactory rejoinder to it. Instead of directly rebutting the Evidentialist thesis, Quong tries to show that even if there is sufficient evidence for the conclusion that citizens are incapable of effectively advancing some goal or interest they have, that does not present a challenge to his thesis about the wrongness of paternalism.

My claim at this stage is very simple. Even if a paternalistic policy is justified, all things considered, it remains true that this sort of policy involves a *prima facie* wrongness. The

wrongness is that it treats citizens as if they cannot make effective decisions about their own good, and thereby diminishes the moral status accorded to citizens. Even if we decide in the end that the benefits of a policy are sufficiently great that the policy should be implemented, it will nevertheless be true that we must weigh these benefits against the substantial cost of denigrating citizens' moral status. In order to rebut my position about the *prima facie* wrongness of paternalism, a critic would need to insist that paternalistic policies present *no threat at all* to citizens' moral status (Quong 2010, pp.102-103).

A plausible interpretation of Quong's rejoinder above is that even if all the relevant evidence makes a negative judgment about citizens' abilities all-things-considered justified, still there is a *residual* wrongness with rendering a negative judgment that is never entirely cancelled out by the all-things-considered justifiability of the negative judgment. This interpretation is strongly implied by Quong's above claim that, "Even if we decide in the end that the benefits of a policy are sufficiently great that the policy should be implemented, *it will nevertheless be true that we must weigh these benefits against the substantial cost of denigrating citizens' moral status*" (emphasis mine) (Quong 2010, p. 101).

But this interpretation is incompatible with Quong's claim that rendering a negative judgment about others' abilities is wrong merely in a *prima facie* sense. A *prima facie* wrongness is a type of wrongness that may be cancelled out upon close scrutiny of the available evidence. This means that if there is sufficient evidence for a negative judgment, then the *prima facie* wrongness of making a negative judgment about others is simply cancelled out, eliminated altogether. So in order to accommodate the suggested interpretation, Quong's view would have to be amended to one that argues that there is a *pro tanto* wrongness with rendering negative judgments about others' abilities which may sometimes be outweighed by sufficient evidence, but is never entirely cancelled out by it.⁵

⁵ In fact, Quong claims that although he endorses the stronger view that paternalism is *pro tanto* wrong, he is committed to defending only the weaker claim that paternalism is *prima facie* wrong in the text considered here. See footnote 66, (Quong 2010, p. 100).

This amended view, however, commits us to the controversial and possibly incoherent view that (1) although we should epistemically believe that the other is inept if all the relevant evidence indicates that, (2) still, we should not believe that the other is inept because it is morally wrong.⁶ But accepting both (1) and (2) is an extremely difficult, if not an impossible task, since it demands us to believe two contradictory things at once. The conflict I identify here cannot be dissolved by pointing out that respect is consistent with holding negative beliefs about another person's capacity without outwardly expressing them. This is because Quong's claim is that paternalistic action is disrespectful because it is *motivated* by a negative judgment about others' capacities, not because it externally *expresses* an insulting message to others (Quong 2010, p. 80). This means that on Quong's view, your behavior could be paternalistic and therefore disrespectful, even if it does not outwardly express contempt for another's capacity. In other words, the wrong-making feature of paternalism is the *belief* that motivates it, not the action itself.⁷ Thus, there is a real tension in Quong's view: on Quong's view, we should both believe that someone is inept if there is enough evidence to support it, but also refrain from believing that because it's disrespectful.

Still, one might argue that there is no real tension in Quong's view because moral considerations of respect raise the threshold of evidence that's sufficient for the belief that others are inept at running their own lives (Fritz 2019; Guerrero 2007; Pace 2011; Hojlund 2021).⁸ According to this "moral encroachment view," you need more evidence for *p* to justifiably

⁶ For a similar point, see (Enoch and Spectre 2021, pp.12-13).

⁷ For a similar interpretation of Quong, see Enoch 2016. Note that some might argue that Quong is concerned with the conjunction of belief and action, not just the former. But even if this were true, it's clear that Quong is primarily concerned with belief, not action. Otherwise, Quong would have argued that the wrongness of paternalism is explained by the *act* of taking away options, withholding information, or some such – not by the *judgment* that motivates those behaviors. In fact, this is why Quong calls his account, the *judgmental definition* of paternalism (see Quong 2010, p. 80).

⁸ Enoch also considers and rejects a similar moral encroachment-based defense of Quong's claim. But my criticism of this defense is quite different from Enoch's.

believe that p , in case believing p about someone entails some sort of moral risk. In the case of paternalistic action, the moral risk entailed by believing that a person is incapable of advancing her own interests consists of denying that person's equal moral status. Thus, on the moral encroachment view, you need more evidence than other situations (where no similar moral risks are involved) in order to justifiably form the belief that someone is incapable of advancing her own interests or goals reliably.

But although it's plausible to think that moral considerations should affect the *attitude or action* you should take given a certain body of evidence concerning another person's ability to advance her own goals, it's unclear why these moral considerations should affect the epistemic justifiability of your *belief* about that person's ability. Regardless of which prudential or moral considerations are at stake, it seems that the same body of evidence for p should provide support for the same degrees of belief in p .⁹ Consider the following pair of cases.

- (1) **Gas Stove High Stakes:** Suppose p consists of the belief that you have turned off the gas stove before you left the house for work. You seem to remember (but not very clearly) that you have turned off the stove. You know that if p is false, your entire 30-stories apartment building could burn down along with your precious dog and the 300 or so residents who occupy the building. You decide to believe that you have turned off the stove, and hurry on your way to work because you don't want to be late.
- (2) **Gas Stove Low Stakes:** Suppose p consists of the belief that you have turned off the gas stove before you left the house for work. In this case, you are living in an isolated country house and you have no pets or anyone else in or around your house. You know that if p is false, you do not risk endangering the lives of anyone else (including pets), although your house might burn down. You decide to believe that you have turned off the stove, and hurry on your way to work because you don't want to be late.

Now it seems that although there are clear differences in the moral and practical stakes involved in the two cases, the available evidence concerning p supports the same degrees of belief in p in

⁹ In this paper, I take "credence" (degrees of belief) and "belief" to be similar concepts, with the only difference being that belief in p is equal to maximum credence in p .

both the high-stakes and the low-stakes cases.¹⁰ Just because there are lower stakes involved in case (2), it does not mean that you have more grounds for believing that you have turned off the stove in case (2), provided that the quality of your memory in the two cases is the same. This means that the sheer difference in the moral or practical stakes does not change how much credence (or degrees of belief) you are justified in assigning to p , given the evidence you currently have on hand.

Nevertheless, it would be practically rational (or similarly, reasonable, wise, prudent, etc.) to be more concerned about going back and checking the stove in the high-stakes case than the low-stakes case because you have more to lose if p is false in the high-stakes case. But notice that this has to do with the *actions* or *attitudes* it's rational for you to take before you decide to believe that p ; this does not have any necessary connection to the justifiability of *believing that p itself* (or the justifiability of forming a certain degree of belief in p), which is fixed by the degree of support that the existing evidence provides. What changes the epistemic justifiability of belief in p is solely the evidence concerning p : how vivid is your memory of turning off the stove, did you smell anything like gas in the house before you left, do you remember hearing a hissing sound, etc.¹¹

¹⁰ Fritz considers a similar pair of low-stakes and high-stakes cases involving a parked car: one (low-stakes) in which you only get a written warning, the other (high-stakes) in which an enraged police officer will kill five innocents, in case you have parked the car illegally. Fritz argues that, holding the available evidence constant, you may not know that p in the high-stakes case even if you do know that p in the low-stakes case just because of the differences in the moral stakes involved (Fritz 2019, pp. 3052-3053). For a similar pairwise example, see Stanley 2005, pp. 3-5.

¹¹ Some might argue that my argument is directed against a strawman because most moral encroachment theorists (with the exception of *radical* moral encroachment theorists) agree with evidentialists that only evidence directly affects the epistemic justifiability of belief (see Bolinger 2020, p. 20). But this objection does nothing to undermine what I have argued so far. The evidentialists I am discussing here – namely, Enoch and the evidentialists he's drawing on – treat “evidence” for justified belief in p as something determined by the observer's sensory experience of the outside world, unconstrained by practical or moral considerations (see Conee and Feldman 2004, p. 84). But Quong's conception of “evidence” for justified belief in another's incapacity (call it p) is constrained by moral considerations of equal respect (where respect constrains justified belief in p by raising the threshold of sufficient

The upshot is that while the seriousness of the moral or practical stakes involved might affect the *attitude or action* it's rational for you to take before deciding to believe that *p*, it should not affect the degrees of *belief* (or levels of credence) that it's epistemically rational to assign to *p*, given a fixed body of evidence concerning *p*. In the case of paternalism, this means that although the moral risk of disrespect might make it reasonable for you to double-check the relevant facts before deciding to believe that someone is weak-willed or foolish, or to express regret after forming a negative belief about her decision-making ability,¹² the existing evidence fixes how much credence you should give to the claim that she is good at advancing her own interests. Thus, the attempt to make sense of the amended version of Quong's response through the lens of the moral encroachment thesis fails.

So it seems that neither Quong's rejoinder to the Evidentialist thesis as it is presented (viz. paternalism is wrong only in a *prima facie* sense) nor the amended version of his response (viz. paternalism is wrong in a *pro tanto* sense) holds up. This suggests that, all other things being equal, forming a negative judgment about others' capacities to promote their own interests is not offensive to their equal moral status, if all the relevant evidence supports a negative judgment.

2. Two Kinds of Evidence: Interpretative vs. Non-interpretative

But what I have argued so far says nothing about what "evidence" for justified belief in other people's lack of agential capacity consists of. According to a prominent version of evidentialism, you have sufficient evidence for believing that *p* in case your sensory experience

evidence for *p*), not just by facts about the agent's behavior. So Quong and evidentialists have different conceptions of which factors affect the evidentiary threshold for epistemically justified belief in another's (in)capacity.

¹² For a similar claim about regret and moral encroachment, see Rapstine 2021.

and mental states support the belief that p , all things considered.¹³ Another way to put this point is that you are epistemically justified in believing that p if your sensory experience and mental states indicate that p is true, all things considered. Following this notion of evidence, Enoch suggests that the evidence for believing that you are the smartest person in your department is analogous to the purely sensory evidence for believing that you are the tallest person in your department. The idea seems to be that if your sensory experience and mental states indicate that you are the smartest member of your department, then you are justified in believing that you are the smartest member, just as you would be justified in believing you are the tallest person around if your sensory experience indicates that.

Having taken into account biases and attempts to compensate for them, and given that I am going to form a belief regarding my comparative intelligence, and that if it turns out I'm the smartest around, I'm not going to say so, or act on this belief, indeed I will go to great lengths to forget that this is so – given all this, if the evidence strongly indicates that I'm the smartest person around, isn't that what I should believe?... Just as I should believe that I'm the tallest person around when this is what the evidence (sufficiently strongly) indicates, I should believe that I'm the smartest person around when *this* is what the evidence (sufficiently strongly) indicates (Enoch 2016, pp. 25-26).

The problem with this response though, is that Enoch ignores an important distinction between the different kinds of questions that are at stake in the examples he gives. While Enoch may be right about the general Evidentialist thesis that the relevant evidence should determine what we should believe about other people's capacities to advance their own interests, his response above collapses the distinction between (1) *non-interpretative questions*, such as questions about how tall someone is, and (2) *interpretative questions*, such as questions about whether someone is smart, which includes the capacity to advance one's own goals and interests effectively.¹⁴

¹³ Conee and Feldman 2004, pp. 83-84.

¹⁴ I take being "smart" to refer broadly to the capacity to register facts well and make sound judgments on the basis of those facts. So a smart person, on this view, is good at judging which goals are conducive to her own interests, as well as good at identifying the right means to her goals.

Interpretative questions are questions which require interpreting a conscious, ratiocinating agent's reasons for her actions or beliefs, whereas *non-interpretative questions* are questions which do not require such interpreting. For example, non-interpretative questions include questions about someone's height, the color of the doormat, or why the ice is melting in the North pole. As these examples testify, observable facts constitute direct, self-explanatory evidence for belief in a proposition. If I am six feet four, and everyone else in my department is below six feet, then of course, I am the tallest person in my department beyond doubt. In this case, there is no need to go through any process of interpreting the reasons why an intentional agent did this or that; simply registering the perceptual difference in our heights through our senses is sufficient to justifiably form a belief about our relative heights.

By contrast, interpretative questions include questions about why someone did this rather than that, questions about whether what they did was justifiable, or whether someone is truly good at advancing her own interests – which are all questions that require understanding the agent's *reasons* for action or belief.¹⁵ This is because, in order to answer such interpretative questions, the “evidence” for a given belief must be based on an interpretation of the reasons in light of which the agent did this or that. For example, in order to determine whether an agent is capable of effectively advancing her own interests, we need to determine whether the agent has sufficiently good reasons for her action. If she does, then she should be judged competent at advancing her own interests, no matter how bizarre or counterproductive her action might appear

¹⁵ For a similar distinction, see Philip Pettit's distinction between “programming explanation” and “interpretative (hermeneutic) explanation” (Pettit 2003, pp. 177-191). Programming explanation refers to a causal explanation of regularities of events, things, or behavior that does not refer to the agent's reasons, whereas interpretative explanation refers to an explanation of the agent's behavior by reference to the agent's own reasons for engaging in that behavior.

on the outside. This shows that in order to determine whether an agent is competent at promoting her own interests, we need to first interpret (or *understand*) the reasons in light of which she acts.

Collapsing the distinction between non-interpretative and interpretative kinds of questions, as Enoch does, obscures an important difference between the two kinds of questions which explains why questions concerning other people's abilities to advance their own interests must be answered through a process involving an interpretation of people's first-personal reasons for action that is not required to answer questions such as questions about people's height differences.

Now one might argue that the interpretative vs. non-interpretative distinction poses no threat to evidentialism, since any plausible version of evidentialism would allow that different kinds of information-gathering practices are required for different kinds of beliefs. But even if Enoch (and the evidentialists he draws upon) might be willing to concede this general point, Enoch does not explicitly recognize the need to understand the agent's reasons for action before making judgments about their intellectual or rational capacities. In fact, by drawing an analogy between the agent's intellectual capacity and the agent's height, Enoch strongly implies that evidence concerning the agent's intellectual capacity can be directly read off of the agent's behavior or other external indicators (e.g. record of accomplishments) just like evidence concerning the agent's height.

But more importantly, the specific type of evidentialism that Enoch draws upon does not appear consistent with the assumption that different information-gathering practices are required for different kinds of beliefs, because it explicitly denies that the justifiability of an epistemic attitude depends upon the "cognitive processes or information-gathering practices that led to the attitude" (Conee and Feldman 2004, p. 84; see also Enoch 2006, p. 26). According to the

evidentialist thesis that Enoch draws on, how responsibly one gathers evidence and deliberates about it has no bearing on the justifiability of one's belief (Conce and Feldman 2004, pp.89-90). But surely, if you are culpably negligent about gathering and evaluating the relevant evidence, such as not taking seriously another person's reasons for action before judging that person to be incapable of advancing her own interests, then your judgment of that person's ability should be considered epistemically unjustified (for a similar claim, see Kornblith 1983). The problem here is not that you are morally blameworthy for not doing what you should to arrive at true beliefs; rather, the problem is that your belief is epistemically unjustified because it is the outcome of a faulty cognitive process. So it seems that the evidentialist thesis Enoch's argument relies on cannot accommodate the interpretative vs. non-interpretative distinction.

The upshot is that in order to draw a warranted conclusion about a person's ability to advance her own interests from facts about her behavior, we need to understand her reasons for behaving that way. But as I will explain in the context of the example below, understanding another person's reasons requires something other than simply observing the other's behavior.

Consider women in certain parts of South Asia who deliberately choose to eat less than their male relatives to the point of dangerously malnourishing themselves.¹⁶ These women's alleged reason for their seemingly perverse behavior is to give more food to their male relatives who generally occupy positions of power and esteem. Suppose that Kiki is a female member of this community and, in line with her community's norms, willingly chooses to eat less than her fair share so that her male relatives can have more to eat, even if that is bound to make her dangerously malnourished. Although Kiki is very much aware of the adverse health effects of her behavior, she still persists.

¹⁶ Miller 1997; Papanek 1990, pp.162-185; Khader 2011, pp. 107-109.

In this case, there are three distinct conclusions an evidentialist might make about Kiki's capacity to advance her own interests. First, it is open to the evidentialist to conclude that Kiki is capable of effectively advancing her overall interests on the grounds that *for Kiki*, the benefits she derives from giving up her share of food to her male relatives outweigh the costs of malnourishment. Second, it is open to the evidentialist to conclude that Kiki is *not* capable of advancing her own interests effectively, because she is apparently doing something harmful to herself. Finally, the evidentialist can suspend her judgment altogether, citing insufficient evidence for any specific judgment about Kiki's ability to promote her interests.

The problem is that although all three conclusions might seem plausible, there are no sufficient grounds for endorsing any of them. The first view is implausible, because it mistakenly assumes that every individual is always infallible about their own interests. As behavioral economists and psychologists have pointed out, no matter how competent and rational a person may be, everyone is more or less vulnerable to cognitive errors and blindspots. So it's question-begging to assume that whatever Kiki does, that must be the best course of action for her.

The second view, by contrast, mistakenly assumes that Kiki *must* be incapable of making sound judgments about her own interests just because she has grown up in an oppressive environment. Although the fact that Kiki has grown up in oppressive circumstances might increase the probability that Kiki has developed an impaired capacity to make sound judgments about her own interests, that is not by itself sufficient grounds for concluding that Kiki's capacity to reason about her interests is impaired.¹⁷ The third conclusion might appear to be most plausible, but this conclusion too, lacks support since concluding that there is insufficient

¹⁷ There is a broad consensus within contemporary feminist thought that it is both unjustifiable and disrespectful to assume that agents living in oppressive circumstances cannot exercise rational agency. See Baber 2007, p. 118; Narayan 2002, p. 418; Christman 2004, p.152; Westlund 2009, p. 29.

evidence for a positive or negative judgment about the agent's ability is *itself* a specific judgment about what the evidence indicates (or rather, fails to indicate). But without interpreting and understanding the reasons in light of which Kiki acts, we have no grounds for concluding that there's insufficient evidence of Kiki's (in)capacity. No matter what the evidentialist concludes about Kiki's ability to advance her own interests then, the conclusion would seem to be unwarranted.

To determine which conclusion about Kiki's capacity is warranted then, we need to understand Kiki's reasons for persisting in what appears to be self-destructive behavior, and assess whether those reasons truly provide rational grounds for her behavior. But notice that simply observing Kiki's behavior is insufficient to understand Kiki's reasons for her behavior, since, to most of those outside of Kiki's culture, Kiki's behavior would obviously appear bizarre and self-destructive to the point of irrationality. Just by observing Kiki's behavior then, we can only judge Kiki as deeply deluded or mistaken about what she is doing. But concluding that Kiki must be mistaken about what she's doing based on our unilateral observation of Kiki's behavior amounts to ignoring Kiki's reasons, rather than taking her reasons seriously as considerations intended to rationally explain why she thinks it's good or right to act as she does.

So contrary to what Enoch claims, there does seem to be a significant difference between questions about people's height differences and questions about an agent's ability to advance her own interests or goals effectively. Unlike questions about who is the tallest person in one's department, questions about an agent's ability to advance her own interests can only be settled on the basis of an interpretation of the agent's reasons for action, because perceived facts about the agent's behavior do not themselves provide evidence for any definite conclusion about the agent's ability to advance her own interests. By contrast, perceived facts about the height

differences between oneself and other members of one's department do provide sufficient evidence for (or against) the conclusion that one is the tallest person in one's department.

Hence Enoch's evidentialist view must be revised so that it can accommodate the relevant distinction between *interpretative* and *non-interpretative* questions.

3. The Case for interpreting Evidence of Incapacity through Discursive Exchange

Now one might think that Enoch's evidentialist view can easily accommodate the problem I have pointed out above. Isn't it sufficient to simply ask Kiki what her reason is, and then determine whether that reason provides sufficient grounds for judging that Kiki is capable of effectively promoting her own interests? But contrary to what this over-simplified view suggests, understanding other people's reasons requires going beyond merely knowing *what* caused them to act in a particular way. Instead, understanding others' reasons requires making sense of *why* they were motivated to see those considerations as "good" reasons for acting that way. And the latter, as I will explain, requires not just asking others to give an account of their reasons and accepting that account whatever it is, but rather working towards a rationally motivated consensus as to why their reasons should be accepted as sufficient grounds for their judgment.

Recall that a "reason," by definition, refers to some consideration that is intended to provide a justification (or grounds) for an action or belief.¹⁸ So for instance, when you give reasons to others for expressing anger and resentment at your friend's dishonesty, you expect others to comprehend why any sound-minded person in your shoes would consider your friend's dishonest behavior sufficient grounds for getting angry at them— rather than simply register the fact that

¹⁸ See Scanlon 2000, pp. 18-19; Dancy 2002, ch.1.

your friend's behavior was causally effective in making you angry. To put this point in general terms, understanding another person's reasons involves not just knowing *what* considerations that person takes as reasons sufficient for her action (or belief), but also knowing *why* that person, who is presumably just as rational and clear-headed as anyone else, believes that those considerations are fit to be taken as a "good" justification for her action (or belief).¹⁹

But in order to know why Kiki believes that her reason is a *good* reason for action, you must clarify the implicit standard of validity that Kiki (and the fellow members of her community) is relying on when she makes her judgment, so that relative to that standard, you can make sense of why Kiki felt justified in putting forth her reason as *good* grounds for her action in her situation. In Jürgen Habermas's terms, you must "rationally interpret," or provide a rational account of the implicit rules that govern what's right or wrong, appropriate or inappropriate to do in Kiki's situation (Habermas 1990, pp. 30-32). Habermas explains the connection between understanding reasons and rational interpretation in the following.²⁰

The interpreter would not have understood what a "reason" is if he did not reconstruct it with its claim to provide grounds; that is, if he did not give it a rational interpretation in Max Weber's sense. The *description* of reasons demands *ep ipso* an evaluation, even when the one providing the description feels that he is not at the moment in a position to judge their soundness. One can understand reasons only to the extent that one knows *why* they are or are not unsound, or why in a given case a decision as to whether reasons are good or bad is not (yet) possible. An interpreter cannot, therefore, interpret expressions connected through criticizable validity claims with a potential of reasons (and thus represent knowledge) without taking a position on them. And he cannot take a position without applying his *own* standards of judgment, at any rate standards that he has made his own... There is then a *fundamental connection between understanding communicative actions and constructing rational interpretations* (Habermas 1985, pp. 115-116).

¹⁹ For a similar distinction, see Pettit's distinction between "normalizing explanation" and "interpretative (hermeneutic) explanation" (Pettit 2003, pp. 182-190). Normalizing explanation explains the agent's behavior just by reference to external rules that are necessary for the agent to successfully perform some function/goal that she is designed to perform, whereas the latter refers to an explanation of the agent's behavior by reference to the agent's own reasoning about the relative value of her options and the ends she takes to be right or appropriate. See also *supra* footnote 26.

²⁰ See also Pederson 2008. For a similar discussion of rational interpretation in the philosophy of science, see Lakatos 1970.

As Habermas notes above, because your rational interpretation of Kiki's reasons inevitably depends on *your* own standard of judgment, you cannot avoid evaluating Kiki's reasons as right or wrong, good or bad relative to your own standards, in the course of making sense of her reasons.²¹ An example of such a rational explanation of Kiki's reasons is Amartya Sen's hypothesis that women like Kiki were most likely induced to accept the misogynistic norms of their community due to the diminished sense of self-worth they have internalized from prolonged exposure to unequal treatment, social conditioning, and coercion (Sen 1990, pp. 123-149). According to Sen's type of explanation, it would be justified to render a negative judgment of Kiki's capacity to advance her own interests, because Kiki's false sense of low self-worth is hindering her from making an accurate judgment about her own interests.

The important point is that if there are no grounds for doubting that Kiki is a moral equal who can competently exercise her rational agency, then there are no grounds for privileging Sen's type of explanation and its associated standard of judgment over Kiki's own (Habermas 1985, pp. 118-119; Habermas 1990, p.30).²² After all, Kiki might have a totally different account of why she came to see her reasons as "good" reasons for giving up her share of food to her male relatives, and by *her standards*, her reasons might be perfectly good reasons for acting as she does. To be sure, this does not mean that Kiki's own standard for judging the (un)soundness of reasons should be accepted as infallible, since Kiki's own standard is bound to be embedded in taken-for-granted cultural traditions and assumptions which must themselves be exposed to public review and criticism. What follows from this is that neither Kiki's own standard nor Sen's standard can be assumed as correct from the get-go.

²¹ See *supra* footnote 8.

²² The presumption in favor of equal moral agency is not only a feature of relational egalitarianism, but more generally a feature of the liberal tradition.

In order to sort out whose standard of reason judgment is correct then, we need to test each standard through a decision procedure that does not arbitrarily privilege anyone's particular viewpoint or conception of interests. One such decision procedure is hypothetical idealization of the sort that characterizes the contractualist theories of Rawls and T.M. Scanlon (Rawls 1971; Scanlon 2000). According to the hypothetical idealization procedure, we decide which standard of reason judgment is best by asking ourselves, "Which standard of judgment about reasons would a perfectly rational agent accept?" But as it turns out, idealization cannot uniquely determine the standard for judging the soundness of reasons because there are many different standards that an ideally rational agent could hypothetically endorse.

For example, an ideally rational agent such as the fictitious agent in Rawls's Original Position might (1) endorse Sen's standard and conclude that Kiki's choice is based on irrational preferences that are the result of psychological adaptation to oppressive circumstances, *or* (2) endorse a different standard and judge that Kiki's choice is a rational strategy in a community where the approval of one's fellow community members (especially, the male members) is considered crucial for the social bases of self-respect. But since hypothetical idealization does not contain or imply any independent criteria by which we can decide whether Kiki has been manipulated or not, we cannot decide whether Sen is right that Kiki has been deluded or whether Kiki has valid reasons of her own through hypothetical idealization alone. In fact, given that an ideally rational agent is typically only required to meet the criteria of procedural (or instrumental) rationality such as internal consistency of preferences, an ideally rational agent might still endorse values that are the result of manipulation or coercion.²³

²³ As Joel Feinberg puts it, "The role of practical reason for Rawls is much the same as for Hume – to serve a person's antecedently given wants, and select efficacious means to his antecedently given ends" (Feinberg 1986, pp. 110-111).

Thus, in order to determine whose standard about reason judgments is valid, we need a decision procedure that allows real persons with their existing preferences and assumptions to idealize from their own standpoint and test their idealized judgments against each other, so that their existing preferences and assumptions themselves could be exposed to mutual critique. The only decision procedure that meets this description is discursive exchange between the idealizer(s) and the subject(s) of the idealization.

To be sure, discursive exchange also involves idealizing since it involves each individual constructing their own idealized account of the agent's interests, and defending their account with rational arguments against others' criticism. But unlike solitary idealization, discursive exchange is a *collective* idealization that requires each person (including Kiki herself²⁴) to put her own idealized account to the test by exposing it to mutual criticism. Therefore, unlike solitary idealization, discursive exchange with others can filter out "bad" idealizations that are merely projections of one's own prejudices or cognitive errors.²⁵ If the arguments so far are correct, then it follows that Kiki's capacity to effectively advance her own interests should also be assessed through a discursive exchange between Kiki and all others concerned to reach a warranted judgment about her capacity.

Now Enoch (and others who agree with Enoch) might object that, setting aside cases like Kiki's case where the agent might be wrong about her *judgment* about which end or goal is truly in her best interests, in cases where the agent repeatedly fails to *act* in ways that are consistent with her own judgment about her goals (due to say, weakness of will), shouldn't we take the

²⁴ Note that Kiki's rational explanation of her own action could be mistaken, just like anyone else's explanation.

²⁵ See Vallier 2019, p.1117; Habermas 1990, p.67. Some might be worried about the familiar empirical concerns about deliberation with others, such as power inequality or lack of consensus. I set aside these concerns here, because they are subjects for empirical research on the conditions for successful deliberation. In this article, I only make the theoretical point that no one can insist on the correctness of their standard of judgment about reasons before rationally arguing it out with the disagreeing party. For a recent empirical study on the usefulness of conversational exchange and learning, see Broockman and Kalla 2020.

agent's repeated failure to act in the right way as sufficient evidence of her incapacity? More specifically, isn't a past record of the agent's repeated failure to act in her own best interests analogous to facts about one's height or weight in directly providing evidence for the conclusion that the agent is incapable of effectively advancing her own ends? But it turns out that in the case of failures to *act* rationally too, understanding the agent's reason through discursive exchange is also required.

Consider an analogue of Enoch's hypothetical case where I am considering a friend's request to borrow £50 to buy flowers for Mother's day. Suppose that this friend has a past record of using her money to buy addictive drugs, which even she admits is bad for herself. But suppose that my friend assures me that *this time* she will spend her money on buying flowers for her mother and nothing else. In this case, an evidentialist in Enoch's camp would argue that regardless of what my friend says, the past record of her behavior is sufficient evidence for concluding that she will misuse the £50 this time again if I lend it to her, and that therefore, I should not lend her the money.

But this view overlooks the fact that a past record of my friend's behavior is not the *only* evidence that is relevant to judging how likely or capable my friend will be of resisting the temptation to spend the £50 on drugs this time. In addition to the past record of my friend's behavior, another relevant piece of evidence is how strong or urgent my friend's reason for wanting to borrow £50 is.

At first glance, one might wonder why the agent's reasons for wanting to do X should affect one's judgment of how likely or capable the agent would be of doing X. Even if the agent has very weighty reasons to use her money for a good purpose in this situation, she might be just as tempted to waste her money in this situation as in other situations (or so, one might think). So

one might think that regardless of whether the agent has weighty reasons to X in this situation compared to other situations, the agent will be just as likely not to X in this situation as in other situations.

But notice that this objection is valid only when the agent is unaware of her reasons to X, or is somehow incapable of properly appreciating her reasons to X.²⁶ *If the agent is appropriately aware of her reasons to X*, then there is generally a positive correlation between the strength of the agent's reasons to X and the likelihood or capability of the agent doing X. This is because the strength of the agent's reasons for doing X are, as a matter of psychological fact and as a matter of rational requirement, proportional to the strength of the agent's motivation to do X. When we think we have especially strong, urgent reasons to X, we tend to be more motivated to X than in normal circumstances. Indeed, it would be irrational to be less motivated or equally motivated to X when we have especially strong reasons to X than in normal circumstances (or conversely, strongly motivated to X when we have relatively weak reasons to X).

For example, if my doctor warns me I have a serious lung disease and is in imminent danger of dying if I do not quit smoking right now, I would be more motivated to quit smoking than if my doctor simply advised me to quit smoking for my general health and well-being. In fact, I would be irrational if I were not more motivated to quit smoking in the face of an imminent threat of death than in normal circumstances. Hence, the stronger or more urgent my reasons for doing X is, the more I would be (and should be) motivated to X.

²⁶ Weakness of will generally does not impair one's epistemic awareness or sensitivity to one's reasons. Rather, weakness of will usually works by weakening or undermining the connection between (1) one's awareness of one's reasons, and (2) one's ability to act in accordance with the reasons one believes one has. So even if one is keenly aware of one's reasons to quit using drugs, for example, one might keep on using because one's ability to carry out one's belief into action has been diminished due to addiction. However, recent studies in psychology and philosophy show that addiction does not necessarily hijack the brain so that one cannot carry out one's belief into action altogether (Pickard 2018; Heyman 2010). So even if my friend is an addict, it's unwarranted to presume that she is incapable of acting in accordance with her own judgment.

Also, depending on the relative importance of the matter at stake, we tend to exhibit varying degrees of competence at resisting weakness of will. When there are important matters at stake, we tend to be more focused and purposeful and therefore, less likely to give in to temptation. But when trivial matters are at stake, we are more likely to give in to temptation because we have less to lose. Therefore, knowing the strength or urgency of the agent's reasons for doing X does and should make a difference to our judgment of how likely or capable the agent will be in resisting the temptation not to X.

The caveat is that each of us have different standards for judging which reasons are more urgent or more important in a particular situation. Even if my friend truly thinks that buying flowers for Mother's day is a reason that is strong enough to enable her to resist the temptation to waste the £50 on drugs, I might still have difficulties understanding why she thinks so. For instance, I might think that, given her past behavior, buying flowers for Mother's day is not a reason that is strong enough to enable her to resist her temptation to spend the money on drugs. Or, I might think that my friend is underestimating the seriousness of her addiction or otherwise deceiving herself, and think that when push comes to shove, she will succumb to temptation in spite of herself.

In all of these cases, my standard for judging the strength of my friend's reasons diverges from my friend's standard for judging the strength of her own reasons. And in case our standards of judgment about reasons conflict in this manner, we have to sort out through a rational argument, whether she is right to judge that her reason for spending the £50 on flowers for Mother's day is stronger than other reasons for action that have failed on past occasions to motivate her to resist the craving to spend her money on drugs (Pettit and Smith 1996, pp. 430-433). In other words, I have to sort out with my friend through conversational exchange whether

there is something different about her reason for borrowing money *this* time that would make it rational for me to conclude that she will not waste her money on drugs this time.

Now, some might worry that sorting this out with my friend through rational discussion is futile because addicts are likely to lie in order to feed their drug habit. But even if we take into account the possibility that my friend might lie, engaging her in a discussion about why I should believe her is *not* futile because it can serve to reveal whether she is lying. In order to successfully deceive me, my friend must, at the very least, *pretend* to abide by the norms of rational discourse.²⁷ That is, she must lie in a consistent and logical manner in order to successfully deceive me. If my friend cannot defend her claims consistently with existing facts and other claims she makes or has made in the past, then this itself serves as evidence for the conclusion that she is lying. It is in this sense that rational discussion can reveal whether one's interlocuter is lying.

In fact, even if dialogue cannot decisively prove or reveal whether my friend is lying, I cannot presume from the get-go that reasoning with her is futile just because she *might* lie. Addicts also have their own reasons for action, and it would be disrespectful to addicts, and epistemically unwarranted to boot, to presume that they must be lying or are incapable of acting consistently with their own judgments given that addicts are also capable of rationally responding to incentives and disincentives, just like non-addicts (Pickard 2018; Heyman 2010; Tekin 2018).

The upshot is that in cases of both (1) *judgments* concerning people's own interests, and (2) *actions* that people take to advance their own goals, we must decide whether there is sufficient

²⁷ This is what Habermas means when he says that "the instrumental use of language is parasitic" on communicative use of language oriented towards understanding (Habermas 1985, pp. 288-289).

evidence of their incapacity through an unconstrained process of discursive exchange with them aimed at understanding their reasons.

4. Conclusion: A Discursive Approach to Paternalism

If what I have argued so far is correct, then two conclusions follow. One is that in order to determine if there is sufficient evidence that another person is too obtuse or weak-willed to advance her own interests consistently with respect for her status as a moral equal, we need to first interpret what the facts about her behavior mean, by understanding her reasons for behaving that way through a process of discursive exchange with her. The other is that when the outcome of an equal and uncoerced exchange shows that, after all, the agent has been mistaken about her own interests or her ability to act rationally *on terms that the agent herself cannot reasonably deny*²⁸, it is not offensive to her moral status to render a negative judgment about her capacity to promote her own interests because such a negative judgment is akin to her own rational self-assessment. Furthermore, such an outcome establishes a *pro tanto* case (but not an all-things-considered case) in favor of paternalistic interference, since interference in this case is aimed at making up for a deficiency that the agent cannot reasonably deny as her own. Of course, in order to establish an all-things-considered justification for interference, various other conditions must also be satisfied: the benefits of interference must outweigh the costs of interference, the mode of interference must not end up imposing alien values on the agent or violate the agent's basic constitutional rights, and so forth.

²⁸ Notice that what the agent cannot *reasonably deny* in this case is not defined solely in terms of the paternalizer's standards, since it's defined through a mutual exchange of arguments between the paternalizer and the paternalizee.

However, if the outcome of exchange is indeterminate or if there are reasonable grounds for denying that there is sufficient evidence of the agent's incapacity, paternalistic interference seems unjustified. In both cases, there is no sufficient evidence of the agent's incapacity, and so it would be disrespectful and unwarranted to presume in either case that the agent is incapable of advancing her own interests. Indeed, interfering with the agent in such a case is bound to do her more harm than good since, after all, external interference is likely to hinder the agent from pursuing her own conception of the good in the best possible way for herself.

References

- Anderson, Elizabeth. 1999. What is the point of Equality? *Ethics* 109: 287-337.
- Baber, H.E. 2007. Adaptive Preferences. *Social Theory and Practice* 33(1):105-126.
- Bolinger, Renée. 2020. Varieties of Moral Encroachment. *Philosophical Perspectives* 34(1): 5-26.
- Broockman, D. and Kalla J. (2020). Reducing Exclusionary Attitudes through Interpersonal Conversation: Evidence from Three Field Experiments. *American Political Science Review* 114(2): 410-425. doi:10.1017/S0003055419000923.
- Christman, John. 2004. Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves. *Philosophical Studies*. 117(1/2): 143–164.
- Conee, Earl and Feldman Richard. 2004. *Evidentialism: Essays in Epistemology*. Oxford: Clarendon Press.
- Cornell, Nicolas. 2015. A Third Theory of Paternalism. *Michigan Law Review* 113(8): 1295-1336.
- Dancy, Jonathan. 2000. *Practical Reality*. Oxford: Oxford University Press.
- De Marneffe, Peter. 2006. Avoiding Paternalism. *Philosophy and Public Affairs* 34(1): 68-94.
- Dworkin, Gerald. 1972. Paternalism. *The Monist* 56 (1): 64-84.
- Dworkin, Gerald. 1988. Paternalism: Some Second Thoughts. In *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.
- Enoch, David. 2016. What's wrong with paternalism: Autonomy, Belief, and Action. *Proceedings of the Aristotelian Society* 116 (1): 21-48.
- Enoch, David and Levi Spectre. 2021. There is no such thing as Doxastic Wrongdoing. Unpublished manuscript.
- Feinberg, Joel. 1986. *Harm to Self*. Oxford: Oxford University Press.
- Fritz, James. 2019. Moral Encroachment and Reasons of the Wrong Kind. *Philosophical Studies* 177: 3051-3070.
- Guerrero, Alexander. 2007. Don't know, Don't kill: Moral ignorance, Culpability, and Caution. *Philosophical Studies* 136: 59-97.

- Habermas, Jurgen. 1990. *Moral Consciousness and Communicative Action*. Cambridge, MA: MIT Press.
- Habermas, Jurgen. 1985. *The Theory of Communicative Action: Reason and Rationalization of Society, Vol 1*. Boston: Beacon Press.
- Heyman, Gene. 2010. *Addiction: A Disorder of Choice*. Cambridge: Harvard University Press.
- Hojlund, Anne-Sofie Greisen. 2019. What should Relational Egalitarians Believe? *Philosophy, Politics, and Economics*. 21(1): 55-74.
- Khader, Serene. 2011. *Adaptive Preferences and Women's Empowerment*. Oxford: Oxford University Press.
- Kornblith, Hilary. 1983. Justified Belief and Epistemically Responsible Action. *The Philosophical Review* 92(1): 33-48.
- Lakatos, Imre. 1971. The History of Science and its Rational Reconstructions. eds. R.C. Buck and R.S. Cohen. *PSA 1970: Boston Studies in the Philosophy of Science* 8, Dordrecht: Reidel, 91-135.
- Miller B.D. 1997. Social Class, Gender, and Intra-household food allocations to children in South Asia. *Social Science and Medicine* 44(11): 1685-1695.
- Moss, Sarah. 2018. Moral Encroachment. *Proceedings of the Aristotelian Society* 118(2): 177-205. <https://doi.org/10.1093/arisoc/aoy007>
- Narayan, Umar. 2002. Minds of their Own: choices, autonomy, cultural practices, and other women. eds. Antony and Witt. *Minds of One's Own*. Boulder, CO: Westview Press.
- Pace, Michael. 2011. The Epistemic Value of Moral Considerations: Justification, Moral encroachment, and James' "Will to believe." *Noûs* 45(2): 239-268.
- Papanek, Hannah. 1990. To each less than she needs, from each more than she can do: allocations, entitlement, and value. In *Persistent Inequalities: Women and World development*, ed. Irene Tinker. Oxford: Oxford University Press.
- Pederson, Jørgen. 2008. Habermas' Method: Rational Reconstruction. *Philosophy of the Social Sciences* 38(4): 457-485.
- Pettit, Philip. 2003. Three Aspects of Rational Explanation. In *Rules, Reasons, and Norms: Selected Essays*. Oxford: Clarendon Press.
- Pettit, Philip. and Michael Smith. 1996. Freedom in Belief and Desire. *The Journal of Philosophy* 93(9): 429-449.

- Pickard, Hannah. 2018. The Puzzle of Addiction. In *The Routledge Handbook of the Philosophy and Science of Addiction*, eds. Hannah Pickard and Serge Ahmed. New York: Routledge.
- Quong, Jonathan. 2010. *Liberalism without Perfection*. Oxford: Oxford University Press.
- Rapstine, Mica. 2021. Regrettable Beliefs. *Philosophical Studies* 178(7): 2169-2190.
- Scanlon T.M. 2000. *What we Owe to Each other*. Cambridge, MA: Harvard University Press.
- Schemmel, Christian. 2011. Distributive and Relational Equality. *Policy, Philosophy, and Economics*. 11: 123-148.
- Sen, Amartya. 1990. Gender and Cooperative Conflicts. In *Persistent Inequalities: Women and World Development*, ed. Irene Tinker. Oxford: Oxford University Press.
- Shiffrin, Seana. 2000. Paternalism, unconscionability doctrine, and accommodation. *Philosophy and Public Affairs* 29(3): 205-250.
- Tekin, Serife. 2018. Brain Mechanisms and the Disease Model of Addiction. In *The Routledge Handbook of Philosophy and Science of Addiction*. Routledge: London.
- Tsai, George. 2014. Rational Persuasion as Paternalism. *Philosophy and Public Affairs* 42(1): 78-112.
- Vallier, Kevin. 2020. In Defense of Idealization in Public Reason. *Erkenntnis* 85: 1109-1128.
<https://doi.org/10.1007/s10670-018-0067-8>
- Westlund, Andrea. 2009. Rethinking Relational Autonomy. *Hypatia*. 24: 26-49.