

# Paternalism and our Rational Powers

MICHAEL CHOLBI

*California State Polytechnic University, Pomona*  
*mjcholbi@cpp.edu*

According to rational will views of paternalism, the wrongmaking feature of paternalism is that paternalists disregard or fail to respect the rational will of the paternalized, in effect substituting their own presumably superior judgments about what ends the paternalized ought to pursue or how they ought to pursue them. Here I defend a version of the rational will view appealing to three rational powers that constitute rational agency, which I call recognition, discrimination, and satisfaction. By appealing to these powers, my version of the rational will view can rank the wrongfulness of paternalistic acts in terms of (a) the extent to which such acts amount to supplanting the paternalized individual's identity as a rational agent with that of the paternalist, and (b) the degree of mistrust in the paternalized individual's rational agency shown by the paternalistic act. My rational powers account thus provides a more complete account of why paternalism is a powerful, but not decisive or absolute, objection to an act or policy. My rational powers account also provides powerful explanations of why rational suasion deflects charges of paternalism; why consenting to intercessions in one's rational agency negates paternalism; why strong paternalism is generally more objectionable than weak paternalism; and why hard paternalism is generally more objectionable than soft paternalism.

## 1. Paternalism and our Rational Powers

Paternalism involves interfering with another person, against that person's will, in order to benefit that person in some way. This standard characterization nevertheless leaves a philosophically crucial question unaddressed: With what capacities or aspects of a person does a paternalistic act or policy interfere? Mill's famous description of his harm principle (Mill 1859, part I, ¶9) offers two competing answers to that question.

On the one hand, Mill claims that others are not 'warranted, individually or collectively' in 'interfering with' a person's 'liberty of action' unless they do so for purposes of 'self-protection', that is, in order to prevent that person from harming others. Interfering with a person's ability to act for her own good is not morally justified. Taken as an expression of opposition to paternalism, Mill's remarks suggest that paternalism targets a person's liberty, understood as the capacity

to act as one desires. Let us call this the *liberty view* of paternalism's target.

On the other hand, some of Mill's remarks emphasize the wrongfulness of certain means by which we might move others to act in desirable ways. Compulsion, control, and coercion are ruled out, whereas 'remonstrating', 'reasoning', 'persuading', and 'entreating' are permissible means to move others to act. These remarks suggest that the heart of paternalism is not that it interferes with another's liberty as such but that it attempts to influence her behaviour through non-rational causes—that it involves non-rational, or at least less than fully rational, means of influence. The paternalist's target is thus not individual liberty *per se*, but an individual's *rational will* or volition.

Here I will largely set aside the liberty view and concentrate on defending the rational will view of paternalism, according to which paternalism involves wrongful intercession in the operations of a person's rational will. I am not the first to do so; my defence will build on the work of theorists such as Seana Shiffrin (2000), Jonathan Quong (2011, pp. 73-107), and Daniel Groll (2012). Central to their conceptions of paternalism is that paternalists disregard or fail to respect the rational will of the paternalized, in effect substituting their own (presumably superior) judgments about what ends the paternalized ought to pursue or how they ought to pursue them. Paternalistic actors thus attempt to 'take over or control' the rational agency of another (Shiffrin 2000, p. 216). Proponents of the rational will view readily agree with the liberty view that paternalism often interferes with individuals' liberty. Indeed, legal paternalism in particular tends to have this characteristic (Quong 2011, pp. 75-6). However, on the rational will view, interferences with a person's liberty are paternalistic only when (and because) they are also unauthorized intercessions in the operations of their targets' rational will.

My defence of the rational will view proceeds as follows. §1 describes the core commitments of the rational will view and shows how the rational will view elegantly accounts for the role of rational suasion and consent *vis-à-vis* paternalism, as well as accounting for certain otherwise puzzling categories of paternalistic action.

§2 provides the foundation for addressing an issue that adherents of the rational will view have by and large not engaged. That an act or policy is paternalistic should give us moral pause. Yet it would be remarkable if even in cases where enormous benefit could be brought about via paternalistic measures, such measures were impermissible simply by virtue of being paternalistic. Paternalism looks like a

powerful, but not necessarily decisive, reason to reject an act or policy. This position—that paternalism is presumptively, but not necessarily all-things-considered, morally impermissible—has a great deal of pre-theoretical appeal.

The rational will view can allow that paternalism has the good or rightmaking feature of advancing its targets' ends or the ends those targets ought to pursue. But the rational will view faces an unappreciated difficulty in explaining how paternalism can be presumptively, but not necessarily all-things-considered, morally impermissible. Paternalistic acts fall into two conflicting moral kinds: They are objectionable because they are intercessions in a person's rational will, but praiseworthy insofar as they are acts of beneficence. Yet a paternalistic act that instantiates the latter kind to a high degree remains objectionable insofar as it belongs to the former kind. The possibility that a paternalistic act might benefit its target to a large enough degree to render it all-things-considered morally permissible seems to require a way of weighing, however inexact, these two considerations. It might be assumed that such weighing involves only one variable, namely, how much a target is benefitted by a paternalistic intercession. It might be assumed that, where  $X$  and  $Y$  are two paternalistic acts,  $X$  is all-things-considered morally justified if the quantum of benefit  $b$  to  $X$ 's target meets or exceeds some threshold, whereas  $Y$  is not all-things-considered morally justified if the quantum of benefit  $c$  to  $Y$ 's target does not meet that threshold. One of my main aims here is to suggest that this picture of the casuistry of paternalism is too simple, for we cannot assume that  $X$  and  $Y$  should be assigned the same weight by virtue of being paternalistic. Paternalistic intercessions can engage with our rational wills in different ways. On an intuitive level, a person who tries to encourage my frugality by insisting that I not offer gifts on her birthday has engaged with my rational will in less far reaching and less morally objectionable ways than the person who tries to encourage my frugality by disabling the pop-up advertisements on my computer that contribute to my compulsive spending. A more credible casuistry of paternalism thus requires incorporating and weighing two variables, the magnitude of benefit to the target of a paternalistic act and the specific ways in which that act engages with the target's rational will. The rational will view would thus benefit from an elaboration of the different ways in which paternalistic acts engage with their targets' rational wills.

I provide such an elaboration by spelling out the internal architecture of our rational agency. More specifically, our rational agency is

constituted by three rational powers, each of which plays a distinct role in agents' choice and action and each of which can be engaged by the paternalistic actions of others. I argue in §3 that paternalistic intercessions are wrong because and to the extent that they intercede in these rational powers: These powers vary in their centrality to an individual's practical identity, such that the more central a power is to the constitution of an individual's practical identity, the more wrongful are paternalistic interventions that engage that power. §4 offers a second way of conceptualizing paternalism's wrongmaking features in terms of intercessions in our rational powers, namely, that paternalistic actors show varying levels of mistrust in the rational competency of others based on which of these powers such actors intercede in. By grasping the internal architecture of the rational will, we can thus see how the rational will view might normatively weigh the wrongfulness of paternalistic acts against their goodmaking feature (i.e., that, when successful, they aid their targets in securing their ends or ends they ought to seek). My *rational powers account* is thus a version of the rational will view that improves upon it by demystifying the variety of ways in which acts or policies can intercede in the rational wills of others.

I turn to an objection in §5: If we have evidence that a potential target of our paternalistic intercessions is in fact untrustworthy in the exercise of her rational powers, why is it even *pro tanto* wrong to intercede in those powers on her behalf? I defend the notion that respect for rational agency requires deference toward agents even when they do not exercise that agency with full competence or capability.

§6 highlights how my rational powers account explains common philosophical stances about different varieties of paternalism.

## 2. The rational will view and its explanatory advantages

Rational will views of paternalism share two features. First, rational will views require that paternalistic acts involve the substitution of the paternalist's judgment or agency for that of her target. Every account of the nature of paternalism agrees that paternalists act, or try to induce their targets to act, in ways that reflect what the paternalist contends is good (in however broad a sense) for the target to do. As the paternalist sees it, her target is disposed to choose or act either in self-defeating ways, by failing to act in accordance with what the target

antecedently believes to be good or desirable, or in foolish ways, by failing to pursue ends or goods that, if the target were more reflective or knowledgeable, she would pursue. What distinguishes the rational will view is its emphasis on how paternalistic intercessions seek to sever the usual causal relationship between the target's judgment or agency and the target's choices or actions by interpolating the paternalist's own judgment. For adherents of this view, this feature is central to the moral objectionability of paternalism. Seeking to guide another's actions or choices by one's own lights rather than theirs represents an 'intrusion into and insult to [that] person's range of agency' (Shiffrin 2000, p. 218).<sup>1</sup>

Second, rational will views think of paternalism as motivated by the belief that the target of paternalism is, when compared to the paternalistic actor, deficient in some facet of her rational competence or judgment.<sup>2</sup> This second feature is a direct consequence of the first: In order for paternalists to believe that the substitution of their judgment for that of their targets is warranted, they cannot believe, on pain of contradiction, that their own judgment is not superior to their targets' with respect to the matter at hand. Of course, this belief in one's superior competence or judgment need not be global: A may believe herself superior to B with respect to financial decisions but not with respect to dietary ones, and so be inclined to financial paternalism but not dietary paternalism, for instance. Moreover, this need entail neither that the paternalistic actor's judgment is impeccable nor that the target's judgment is incompetent. A may believe that, while she is merely competent with respect to financial decision making, B is a complete financial neophyte who would benefit from her paternalism

<sup>1</sup> As I see it, the intrusiveness of paternalism is normatively fundamental and the insult-iness derivative, that is, paternalism insults others because it intrudes upon others' rational agency, not vice versa. Cornell (2015) appears to disagree, offering an account of paternalism in which its insulting 'expressive content' is fundamental. Others who emphasize that paternalism is insulting or demeaning include Kleinig (1983), Feinberg (1986, pp. 23-4), and Anderson (1999).

<sup>2</sup> Groll (2012, p. 718) presents his version of the rational will view in terms of the target being 'likely to fail to exercise a capacity for sound judgement in the situation at hand'. Quong (2011, pp. 80, 83) argues that paternalism is better described in terms of a 'negative judgement' about the target's abilities than as a judgement that the target's abilities are inferior. His chief reason for this is that paternalistic actions sometimes depend on the belief that their targets lack willpower, rather than poor judgement, and in these cases, the paternalizer need not think that her willpower is superior to her targets. I do not share Quong's assumption that lack of willpower is not an instance of judgement, and while I agree that the paternalizer need not believe his willpower is superior, his judgement is putatively superior in the sense that he believes himself better situated to enable his target to realize her chosen ends.

(whereas C, a financial expert, would be in a position to act paternalistically with respect to B and to A). Furthermore, the rational will view does not entail that the paternalist literally sees her target as childlike or infantile. She merely takes her target to be less competent in her rational judgment or competency than she.<sup>3</sup>

So presented, the rational will view understands the wrongfulness of paternalism as a complex wrong with both comparative and substantive aspects (Quong 2011, pp. 100-2). First, paternalists treat their targets as if they are, with respect to the choices in which the paternalist intercedes, inferior to them. Second, their treating their targets in this way is manifest in intercessions in the operations of their agency or volition. As I see it, neither the comparative nor the substantive aspects of paternalism are more morally basic, and they cannot be disentangled without losing sight of the moral essence of paternalism. Intercessions into other agents' rational volition need not be paternalistic. Lying, arguably, intrudes into others' rational volition but is often not paternalistic. Similarly, there are other ways to treat others as inferior without being paternalistic. But in concert, these two aspects underscore how the moral essence of paternalism is to deny its targets their status as equal rational participants in social life, thereby impugning, as Rawls (2005, p. 19) puts it, their 'capacity to form, to revise, and rationally to pursue a conception of [their] rational advantage or good'.

Furthermore, the rational will view's conception of paternalism's wrongfulness need not exhaust the ways in which paternalistic acts can be wrong. Indeed, one complication in evaluating the moral seriousness of paternalistic acts is that the vast majority of such acts also fall into other morally significant kinds. Most obviously, paternalistic acts are beneficent insofar as they aim at their targets' good. Conversely, many paternalistic acts are also wrongs of another kind. They are acts of violence, promise breaking, coercion, deception, manipulation, intrusions into privacy, etc.<sup>4</sup> As a result, it can prove difficult to distill the distinct contributions that paternalism makes to the all-things-considered moral properties of a paternalistic act. For instance, to what degree is an act of paternalistic coercion wrong because it is coercive and to what degree is it wrong because it is paternalistic? I make no pretence of answering such questions with any precision here. My

<sup>3</sup> De Marneffe (2006), for one, exaggerates Shiffrin's account in this regard.

<sup>4</sup> I hesitate to say that *all* paternalistic acts also fall into other act kinds that are typically wrong. Paternalistic offers, for example, seem to lack any other morally troubling properties.

main objective is to defend the rational will view as the correct account of the bad or wrongmaking feature of paternalism, and it would be far too ambitious to offer a comprehensive casuistry of paternalistic acts. However, I will later appeal to readers' intuitive responses about the moral wrongfulness or objectionability of different paternalistic acts, and there is the very real danger that these responses will be coloured by the fact that these acts also belong to other morally objectionable kinds. I make these remarks to urge readers to focus on the paternalistic features of these acts and to set aside, to the extent possible, other good or badmaking features of these acts.

The rational will view appears to understand paternalism as a psychological kind, as a set of attitudes that paternalists bear toward their subjects. However, this psychological nature of paternalism can be easily misunderstood. That paternalists necessarily have a certain set of attitudes toward their subjects implies neither that paternalists know they have these attitudes nor that they act under descriptions referring to those attitudes. Indeed, in an important sense, paternalism is always well intentioned, inasmuch as it seeks to make its subjects better off. But this is wholly compatible with paternalists failing to recognize how their paternalistic acts rest on judgments of their subjects' relative inferiority or how those acts intercede in their subjects' rational volition. Hence, paternalism is perhaps best understood as an action-theoretical kind, akin to 'cruelty'. Cruelty is most readily characterized by referring to the psychology underlying cruel acts, that is, as the wilful and knowing subjection of another to pain or distress. But characterizing cruelty in this way does not preclude there being unintentional cruelty, accidental cruelty, cruelty done through ignorance, etc. Similarly, I have characterized paternalism psychologically, in terms of the judgments and attitudes of paternalistic actors, even though paternalism, like cruelty, can be accidental, unintentional, done through ignorance, and the like. In this case, it proves perspicuous to characterize the nominative (paternalism, or paternalistic acts) in terms of the adverbial (acting paternalistically).<sup>5</sup> Moreover, if paternalism is an action-theoretical kind, then it is possible to ask relevant moral questions about 'impure' instances of paternalism. Just as we may ask whether, say, unintentional cruelty is as blameworthy as intentional cruelty, we can ask whether unintentional paternalism is as

<sup>5</sup> I take my claim that paternalism is an action-theoretical kind to be in the spirit of the proposal in Grill (2007) that paternalism rests on the 'invocation of certain reasons for certain actions'.

blameworthy as intentional paternalism. I do not address such questions here. I merely point out that understanding paternalism as an action-theoretical kind enables us to differentiate between, and morally interrogate, instances of paternalism whose psychological underpinnings differ from the paradigm instances of paternalism.

Understanding paternalism as an action-theoretical kind also enables the rational will view to account for two sorts of paternalistic acts that might otherwise seem puzzling.

First, the rational will view can allow for paternalistic omissions (Dworkin 2013a, p. 29ff.). The simplest example is a paternalist withholding information from her target for the latter's benefit. Admittedly, such withholding does not interfere with the target's rational deliberation inasmuch as the target will deliberate, and presumably choose and act, just as she would have. This act does not disrupt its target's extant trajectory of rational deliberation. However, all paternalistic reasoning operates with counterfactuals: The paternalist reasons that by interceding in her target's rational agency, the target would (or will) choose or act differently—and better. According to the rational will view, these counterfactuals will sometimes concern what one's target will do if one *omits* some action, thereby ensuring that the target will continue to deliberate, choose, and act as the paternalist expects. But the rational will view recognizes no basis for normatively distinguishing between these counterfactuals and counterfactuals about what one's target will do if one *performs* some action, thereby attempting to guide the target's deliberation, choice, and action in the ways the paternalist believes are better for the target. In either case, the paternalist seeks to regulate her target's rational agency because she believes her own agency superior.

In a similar vein, the rational will view can also explain how offers can be paternalistic. For example, suppose A, believing B is inferior to her with respect to his rational capacities, bribes B to do X because A believes B ought to do X for B's own benefit. This is an instance of paternalism because the reasons A's bribe provides B for doing X do not relate to the reasons B may have had for X prior to the bribe. A has given B *a* reason to act as A believes B ought to act, but assuming that the choice B must make is a typical one, this is the wrong kind of reason, a state-given rather than an object-given reason, to use Derek Parfit's nomenclature (2001). A has in effect made it the case that B ought to do X but not for the reasons that merit B's doing X. Admittedly, A has not circumvented B's rational volition. Yet a condition of A's showing respect for B's rational capacities is that A



rationally persuades B to do X on the basis of X's independent merits. Such paternalistic offers may be all-things-considered justified, but like some instances of rational persuasion, they nevertheless exhibit the disregard for another's rational capacities that morally speaks against paternalism.

Let us conclude this section by noting how the rational will view accounts for two widely acknowledged normative features of paternalism.

First, the rational will view provides a simple explanation of why attempting to influence another's choice or conduct through transparently rational means typically deflects the charge of paternalism (Shffrin 2000, p. 213). For even when one's behaviour has other features commonly associated with paternalism (for example, that one aims to induce another person to act in ways that one perceives to be in their interest), to provide reasons in the hope that doing so will move another to act in a particular way indicates a desire to engage his or her rational will. Providing reasons in the sincere hope of persuading another to act as you believe she ought expresses minimal confidence in the other's rational powers and amounts to giving the individual the opportunity to demonstrate her rational competency or judgment. But if a willingness to see if a putative target of paternalism can be moved by sound reasons, and is therefore trustworthy with respect to her rational competency or judgment, is evidence of non-paternalistic intent,<sup>6</sup> then overtaking another's rational will by substituting one's own competency or judgment for the other's turns out to be essential to paternalism.

We should not, however, tie paternalism too closely to the rejection of rational suasion. As I will detail further in §5, paternalism fails to respect the other person *qua* rational agent. But as George Tsai (2014) has recently argued, the mere fact that one provides reasons to another in an effort to persuade her to act as one believes she ought does not entail that one shows full respect for her rational agency. In providing others with reasons, we do not ignore their rational capacities nor do we directly substitute our judgments for theirs. But we can provide others with reasons in ways that nevertheless fail to honour their capacities for rational volition. We can, through the timing, tone, etc., of our interactions with others provide them with reasons

<sup>6</sup> I say 'evidence' because the willingness to see if rational suasion will move another to act as one believes she ought is nevertheless compatible with lacking respect for her rational agency. One might simply find rational suasion more convenient, etc., than non-rational, paternalistic means of influencing her choices and conduct and so attach no special importance to her rational agency.

whilst nevertheless mistrusting their capacities as rational agents, thus precluding their exercising their decision making capacities in ways that are robust, authentic, and independent. For example, inundating another agent with reasons relevant to a particular choice while subtly pressuring him to make a speedy choice can make it difficult for that agent to deliberate solely by reference to the reasons germane to that choice. Tsai shows that there is more to respecting others' rational capacities—and hence more that is necessary to avoid paternalism—than the provision of reasons. We must allow them to engage fully with those reasons on their own terms, to work out for themselves how such reasons bear on their choices. Only then can others exercise their rational agency competently and authentically.<sup>7</sup> In order that we not be paternalistic in trying to persuade another, we must rationally persuade her (as Aristotle might have said) in the right manner, at the right moment, and in the right spirit. Indeed, when we act paternalistically, it can be said that our acts fail to exhibit the *virtue* of respect.

Secondly, the rational will view offers an elegant explanation as to why another's consent deflects charges of paternalism. When another consents to our treating her in a particular way, we may still harbour the belief that we have better judgment than she. Yet so long as a person's consent to our conduct is minimally rational, our treatment of that person neither fails to engage her will nor amounts to literally substituting our judgment for hers, since her consent provides the judgment that invalidates a charge of paternalism.<sup>8</sup> Consent directly undermines central features of paternalism as the rational will view understands it. The rational will view can thus appeal to a condition or property—that paternalism fails to engage another's rational agency—intrinsic to paternalism itself in order to explain why persuading another or acting with her consent obviates the charge of paternalism.

### 3. Three rational powers

It might appear that when an instance of paternalism is morally justified, this is only because its target is benefitted to a certain degree.

<sup>7</sup> See Christman (2014) for an elaboration of competency and authenticity as conditions of autonomy.

<sup>8</sup> Such consent may not invalidate other charges levelled at such conduct (for example, that it is manipulative, deceptive, etc.). Furthermore, the presence of consent makes it possible to consent to being subject to force, for example, as psychiatric patients do when they consent to Ulysses contracts regarding their treatment while incompetent, without such force being rejected as paternalistic.

That model, I shall suggest here, is too simple to ground an adequate casuistry of paternalism. According to the rational will view, the wrongmaking feature of paternalism is that paternalism amounts to denying equal moral status to another because it intrudes into her rational agency in ways that indicate judgments of inferiority. This wrongmaking feature comes in degrees, I shall argue, because paternalistic acts form a heterogeneous class. Paternalistic acts differ, I shall argue, with respect to which of the rational powers that constitute our rational agency they intercede in. In seeking the rational powers that constitute rational agency, we are seeking the capacities necessary for an individual to choose, intend, and act on reasons whose authoritativeness or reasonableness she recognizes. The rational powers are therefore those capacities needed for practical deliberation and choice to be possible. They are what we possess ‘just in virtue of being an agent’ (Shiffrin 2000, p. 219).

Here I will focus on those rational powers unique to practical, as opposed to theoretical, rational agency. In his discussion of the value of individuality, Mill asserts that those who do not aspire to individuality, those content to let the world choose their plans of life for them, need nothing more than the power of ‘ape-like imitation’. But those seeking to forge their own paths in life must possess certain rational powers:

He who chooses his plan for himself, employs all his faculties. He must use observation to see, reasoning and judgment to foresee, activity to gather materials for decision, discrimination to decide, and when he has decided, firmness and self-control to hold to his deliberate decision. And these qualities he requires and exercises exactly in proportion as the part of his conduct which he determines according to his own judgment and feelings is a large one. (Mill 1859, ch. 3, ¶4)<sup>9</sup>

Here Mill highlights three of the powers necessary for individuals to be practically rational agents.

First, in order for practical deliberation and choice to commence, we first ‘gather materials for decision’. An agent must begin her deliberation from some menu, however small, of options that she perceives as in some respect minimally choiceworthy. We might call these options reasons, considerations, concerns, etc. For ease of exposition, I will call these options *ends*. Not every agent perceives every potential

<sup>9</sup> Quong (2011, p. 81) similarly claims that the relevant abilities about which the paternalist has a negative judgement are her targets’ ‘practical reasoning, willpower, and emotion management’.

end as minimally choiceworthy. This can be due simply to ignorance of it as a possible end. In other cases, an agent can be aware of a prospective end but see nothing intelligibly good about that end, nothing that speaks in favour of it as an object of rational pursuit. Warren Quinn (1993) offers the example of an individual with a desire to turn on every radio she encounters. For most agents, the end served by acting on this desire is, absent further information, simply not intelligible. There is no description of this end under which most agents find its pursuit good in the least. Granted, differences in individual psychology can often explain why one agent finds an end minimally choiceworthy where another does not. And although this capacity to perceive ends as choiceworthy is more a passive power like perception than an active power like choice, it is a rational power in that it involves a sensitivity to those facts that give us reasons to adopt attitudes toward particular objects. Let us call this first rational power, the power to recognize ends as minimally choiceworthy, the power of *recognition*.

Agents with the power of recognition must also be able to choose among these ends and endorse some among them as worthy of being pursued in action. This is a rational power in two senses. First, agents sometimes must weigh competing or incompatible ends in order to engage in rational choice or action. Second, in order for choice to be rational, it must respond to the features of the ends chosen that render those ends choiceworthy. In both of these respects, agents must attend to and attach evaluative significance to various features of ends. Following Mill, let us call this capacity to rationally select, from among those goods one recognizes as minimally desirable or choiceworthy, which goods to endorse and thereby render objects of rational pursuit the power of *discrimination*.

Finally, agents must possess the ‘firmness and self-control’ necessary to put their choices and intentions into action. Action essentially involves pursuing those means, either causal or constitutive, to realize the ends selected via the power of discrimination. Doing so requires the exercise of instrumental rationality, or as I will call it, the power of *satisfaction*.

Let us fix on a set of examples to illustrate how paternalistic intercessions differ with respect to which of these rational powers are engaged.

C, a recovered alcoholic, is aware of the genetic and environmental factors that make it more likely that children of alcoholics will become problem drinkers. Thus, C desires that her adult child, D, refrain from

drinking alcohol. She desires this for D's sake and believes that D's judgment or competency with respect to choosing whether to consume alcohol is weaker than hers. She would very much prefer that D not have to wrestle with alcohol in order to perceive its dangers, as C did. If the above picture of our rational powers is correct, then C has at her disposal several levers that she might pull in order to influence D's rational will in the hope that D will not drink. For reasons that will become apparent shortly, let us confine our attention to paternalistic intercessions in which C lies to or deceives D.

C could try to make it the case that D has no desire at all to drink alcohol. The most decisive way to do so would be for C to eliminate drinking from the menu of ends from which D can choose. Admittedly, it is difficult to imagine how to do this. Perhaps C could undertake a long-term program of 'censorship' directed at D, designed to keep D ignorant of alcohol and its allure. She might dine with D only at restaurants that do not serve alcohol, change the television channel whenever alcohol advertisements are broadcast, interfere with D's personal relationships so that none of D's friends imbibe, etc. If C's censorship campaign succeeds, then drinking will not appear on the menu of minimally choiceworthy ends that D recognizes.

Another way for C to forestall D's desiring to drink alcohol is to 'tilt the scales' against drinking by making it seem relatively less choiceworthy. C could achieve this by wilfully exaggerating the dangers of alcohol to D, thus making D less likely to drink.

Finally, where C is confident that D endorses abstinence as an end, C might take measures to ensure that D prudently pursues that end. C might, for instance, deceive D about whether D's companion, known to be a heavy imbibor, has called to suggest that D join him for an evening at the local watering hole.

The ways in which C might try to prevent D's drinking illustrates how different paternalistic intercessions target different rational powers. When C undertakes the campaign to suppress D's awareness of alcohol and its appeal, C intercedes in D's power of recognition, taking alcohol off the menu of minimally choiceworthy ends that D may choose to pursue. When C exaggerates the dangers of alcohol to D, C intercedes in D's power of discrimination, trying to make it the case that drinking is not an end that D finds sufficiently attractive to pursue, despite D's continuing to recognize alcohol consumption as a minimally choiceworthy end. In deceiving D regarding the invitation to the local watering hole, C intercedes in D's ability to successfully act on his choices and so intercedes in D's power of satisfaction.

These three powers are obviously not sufficient on their own for practical agency. As Mill saw, practical agents must also have other perceptual and cognitive capacities in order to exercise their powers of rational choice. But these powers do constitute the distinctly practical or choice-orientated core of powers that a rational agent has.

#### **4. Rational powers and respect for practical identities**

With this sketch of the three rational powers in hand, I propose that the rational will view of paternalism, according to which paternalism involves substituting one's judgment for another's and thereby failing to engage another's rational will, ought to conceptualize the wrongs of paternalism in terms of which of these rational powers paternalistic acts seek to supplant. In other words, by understanding how various kinds of paternalism attempt to supplant the three powers necessary for rational volition, we can understand how various paternalistic intercessions are illegitimate based on which of these powers they intercede in. More specifically, this account locates the grounds for opposition to paternalism in truths about individual identity. According to this account, paternalism is a condemnation of our competency as rational agents. But it is also a condemnation of who we are—of important elements of our identity as individuals—and this gives us a special reason for opposing paternalism.

As we noted in §1, one challenge associated with judging the wrongfulness of paternalism is that paternalistic acts are very often also acts of another morally objectionable kind. This fact can make specific judgments about the wrongfulness of paternalistic acts problematic. For argumentative purposes however, the three examples provided in the previous section have the advantage that they are also wrongs of the same kind (or similar kinds), lies or acts of deception. Hence, this set of examples should allow us to set aside the other wrongmaking features paternalistic acts typically have and distil the distinctive paternalistic wrongs involved in each.

Our individual identities are not exhausted by our rational wills. But the rational will dominates the practical identity around which a large portion of interpersonal morality is orientated. When others treat us paternalistically, they judge us inadequate or inferior as rational agents and substitute their judgment for ours on our behalf. Others' paternalism thus colonizes our identities as rational wills. In interceding in the rational powers that constitute our rational agency,

paternalists act wrongly because they attempt, perhaps only momentarily, to be us—to operate the levers of our rational agency on our behalf.

But these intercessions colonize our rational wills in different ways, based on which of the rational powers are interceded in.<sup>10</sup> Hence, the precise ways in which paternalists seek to supplant some element of our practical identities varies, and as we might anticipate, the wrongfulness of paternalistic acts is partially determined by how they supplant our practical identities. In fact, the rational powers stand in relations such that intercessions in one power may also effectively supplant others.

As Christine Korsgaard (1996, p. 100) puts it, ‘When you deliberate, it is as if there were something over and above all your desires, something which is you, and which chooses which desire to act on’. This standpoint ‘over and above’ one’s desires is one’s rational agency or will. The rational will—that something ‘over and above’ our particular choices and desires—engages with possible objects of choice in order to fashion plans or agendas for action. Competency as a rational agent, and one’s identity *qua* rational agent, are thus determined by the relation between one’s rational will and the objects of rational choice. Paternalism ‘comes between me and my plans for myself’ (Conly 2013, pp. 77, 190–2). When C intercedes in D’s power of satisfaction by trying to prevent D from joining his companion at the local drinking establishment, C shares D’s agenda but doubts that D can successfully pursue that agenda. Hence, C has interceded in D’s agenda in the weakest way possible, and though D has reason to object to such an intercession, these reasons are relatively weak because one’s power of satisfaction is not a power that is especially definitive of oneself as a distinct practical agent.<sup>11</sup> Given D’s agenda to forego alcohol, the choice of how to satisfy this end is essentially the same choice faced by any agent with the same agenda and equipped with the same resources, constraints, etc. By settling all the normatively important facts and settling on his practical agenda, D has already set

<sup>10</sup> I do not intend that an act of paternalism can only intercede in one of these three powers. No doubt some such acts intercede in more than one of these powers.

<sup>11</sup> This claim comes with an important ‘all other things being equal’ clause: It may be that for some deeply important end of mine (attending the church of my choice), interceding in my power of satisfaction is something I would very strongly object to, perhaps more strongly than I would object to having my power to satisfy some other less important end interfered with (purchasing the brand of pasta of my choice). But I have in mind that different sorts of intercessions in rational volitions are more or less objectionable relative to a given end, not across different ends. All the other claims I make in this section regarding the objectionability of paternalistic acts should be read in a similar, end-relative light.

the parameters under which C's intercession can be successful. So though C intercedes in D's agency by interceding in his power of satisfaction, this intercession is relatively incidental to who D is as a rational agent with a particular history of decision and choice. The rational power C supplants while acting on D's behalf, while not trivial in value, is not especially central to D's identity as a distinct rational agent.

However, when C intercedes in D's power of discrimination by lying to D about alcohol's dangers, C seeks to influence D's understanding of the facts that determine D's choice agenda. With respect to this particular agenda of D's, C acts on D's behalf by seeking to shift the weight of reasons D recognizes in favour of not drinking. Notice that even though C intercedes in the power of discrimination, her intercession (if successful) will also shape how D exercises his power of satisfaction. If D ends up believing C's lies about alcohol's dangers and makes refraining from alcohol consumption his agenda, then D must now exercise his power of satisfaction in order to satisfy this new agenda. Whereas effective exercise of D's power of satisfaction previously demanded he ascertain how to continue to drink, now effective exercise of that power demands he ascertain how not to. The power of discrimination thus encompasses the power of satisfaction in that the exercise of the former determines what counts as an effective exercise of the latter. Hence, intercessions in the former are, in comparison with intercessions in the latter, more thoroughgoing colonizations of one's rational powers, and hence amount to more complete takeovers of one's identity as a rational being.

This colonization of one's identity as a rational being is even more extensive when one intercedes in another's power of recognition. Here a successful intercession determines not just the actual agenda a person will enact, but the scope of possible agendas she may enact. Intercessions in the power of recognition are hence intercessions in an agent's 'meta-agenda'. C's attempted censorship campaign bears the same relation to D's power of satisfaction as did C's lying to D about alcohol's dangers. Now D must determine how to refrain from drinking. But in interceding in D's power of recognition, C also shapes what successful exercises of D's power of discrimination look like. Absent C's intercession in D's power of recognition, D's power of discrimination would be directed at drinking as a possible end. With drinking removed from D's deliberative field, D can safely shape his practical agenda without attending to the merits of alcohol. Thus, while the power of discrimination encompasses the power of satisfaction in that



the exercise of the former determines what constitute successful exercises of the latter, the power of recognition encompasses both of these, determining the conditions for the successful exercise of both powers.<sup>12</sup>

We thus have reason to morally object to different forms of paternalism depending on which of our rational powers are interceded in, because these powers play different roles in determining how we rationally choose and deliberate and so play different roles in constituting us as rational agents. One might nevertheless believe that any of the three intercessions described in the examples above are morally justifiable despite their being paternalistic. However, if I am correct, the justificatory burden a would-be paternalist faces depends in part on which of the three rational powers she intends to intercede in. The reasons needed to justify paternalistic intercessions in the power of recognition must be stronger than the reasons needed to justify paternalistic intercessions in the power of discrimination, which in turn must be stronger than the reasons needed to justify paternalistic intercessions in the power of satisfaction. Put in contractualist terms, we have stronger reasons to endorse a principle disallowing intercessions in the power of recognition than to endorse a principle disallowing intercessions in the power of discrimination (though of course we have reasons to endorse both such principles). Likewise, we have stronger reasons to endorse a principle disallowing intercessions in the power of discrimination than to endorse a principle disallowing intercessions in the power of satisfaction. Put differently: While our practical identities—the goals, values, or concerns we identify with—are no doubt influenced from many directions, the ultimate determination of our practical identities should largely be left up to us. Paternalism is wrong because, and to the degree that, the paternalist regulates the rational powers through which we fashion our identities.

Note that these claims are issued from the point of view of someone targeted for paternalism, i.e., the reasons we have for rejecting

<sup>12</sup> A referee for this journal suggests that recognition and satisfaction resemble one another in that interceding in these powers involves arranging the world in a certain way, whereas interceding in the power of discrimination is more direct in that the intercession aims to give faulty weights to different possible ends. I grant that interceding in the power of discrimination is more *causally* direct than interceding in the other powers. This observation illustrates that interceding in a person's rational powers can involve the modification of the powers themselves (or their operations) or the modification of the objects ('the world') at which those powers are directed. However, this observation does not undermine the claims I make here about the priority relations among the powers with respect to rational deliberation and agency.

paternalistic intercessions in our own choices and actions vary according to which of our rational powers such intercessions engage. For their part, paternalists will presumably intercede in whichever power it is easiest to engage so as to achieve their aims. Hence, even if (for example) a paternalist could benefit her target by interceding in her target's power of discrimination, she may find it easier to achieve that same aim by interceding in the target's power of satisfaction. Indeed, if my claims are correct, a conscientious paternalist will try to minimize her 'colonization' of her target's rational powers by opting to intercede in the least fundamental power compatible with her achieving her beneficent aim.

No matter whether a paternalist engages with the power of recognition, of discrimination, or of satisfaction, the *pro tanto* wrong of paternalism belongs to the same generic kind: She intercedes in those powers so as to lead the agent to choose or act as she believes the agent ought to choose or act, and she does so from the belief that she is more competent than her target with regard to such choices and actions. However, by arraying paternalistic intercessions in terms of which rational powers they engage, the rational powers account provides a richer account of how paternalism is presumptively, but not necessarily, impermissible. Suppose we assume that paternalistic intercessions can be all-things-considered justified if they are likely to augment their targets' welfare (or advance their ends) to some sufficiently high degree. The rational powers account asserts that, for some given quantum of welfare, the fact that a paternalistic act will likely provide that quantum of welfare to its target must be measured against which of the rational powers that act intercedes in. If for example a paternalistic act provides sufficient welfare to its target to justify an intercession in the power of recognition, then it also provides sufficient welfare to justify an intercession in the powers of discrimination or satisfaction (since justifying intercessions in these powers faces less of a burden than justifying intercessions in the power of recognition). The converse, however, will not necessarily hold. An intercession in the power of satisfaction likely to yield a given quantum of welfare to its target may not be great enough to meet the justificatory burdens imposed by the powers of discrimination or recognition. Note that making welfare relevant to the all-things-considered morality of paternalism does not make the rational powers account 'welfarist': Paternalistic acts, just in so far as they are paternalistic, trace their wrongfulness back to which of the rational powers are interceded in, and there is no inherent relationship between these powers and welfare. It is not the case, for

example, that intercessions in the power of recognition necessarily contribute more (or detract more) from their targets' welfare than intercessions in the other powers. So too for any other goods that might be realized via paternalism (improvements in moral virtue, say). The goodmaking features that may play a central role in the all-things-considered justifiability of paternalism are therefore logically and normatively independent of paternalism's wrongmaking feature, that it intercedes in others' rational agency or volition.

And again, we must keep in mind that these features in turn logically interact with other morally salient features of paternalistic acts (that they are deceptive, manipulative, violent, etc.) that determine the all-things-considered moral permissibility of such acts. For instance, an act that interferes with a target's power of satisfaction (tying a person to a chair so that she cannot eat a scrumptious dessert) may, from a broader moral perspective, be more objectionable than an act that achieves the same end by interfering with a target's power of discrimination (misleading the target about the dessert's scrumptiousness, say), despite the latter's being more objectionable, *qua* paternalistic act, than the former. How we intercede in another's rational powers can be as morally significant as what powers we intercede in. But this, in my estimation, is what we should expect from an account of paternalism's wrongness: an explanation of why paternalism is a strong reason against an act or policy, but not one that accounts for every moral qualm we might have about paternalistic acts.

## 5. Paternalism and mistrust

A second way of understanding the wrongmaking feature of paternalism (on my rational powers account) appeals to the varying levels of mistrust that paternalists show when they intercede in a person's rational powers. The rational will view maintains that paternalism's wrongfulness rests on paternalists' manifesting an attitude of superiority with respect to others' rational volition or competency. We desire to be trusted by others, to be seen as rationally competent custodians of our own affairs. Being trusted is a prerequisite for inclusion, equal standing, and the kind of solidarity amidst diversity that characterizes healthy social relations. Conversely, when we fail to trust in one another's rational powers, marginalization, hierarchy, and opacity in our social relations are the likely results.

The three rational powers play distinct roles in our practical deliberation, and as a result, those who intercede in these powers show distinct levels of mistrust in our competency as rational agents. Consider first C's intercession in D's power of recognition, suppressing D's awareness of alcohol altogether so that D cannot recognize alcohol consumption as even minimally choiceworthy. Such an intercession profoundly manipulates the terms in which D can subsequently exercise his rational will. By taking alcohol consumption off the menu of minimally choiceworthy ends D might choose, C has sought to preclude D from selecting this as one of his ends. D ends up ignorant of alcohol's charms, and in this case, D's ignorance of the merits of this end may well be bliss. Still, C has concluded that D's judgment in this regard is sufficiently inferior to hers that it is best not even to risk the prospect that D will choose to drink, and as a result C's paternalistic intercession makes it impossible for D to rationally appraise the merits of drinking. D ends up not choosing to drink, but not thanks to having been exposed to the merits of alcohol consumption. Hence, with respect to alcohol consumption, C has substituted her judgment for D's in the most thorough and wide reaching way possible. By removing alcohol consumption from D's deliberative field altogether, C has determined that D's judgment is so impaired that D ought not to deliberate about alcohol's merits at all. C's beliefs thus reflect a *deep* mistrust of the competency of D's rational will or agency.

In contrast, by interceding in D's power of discrimination, C does not remove alcohol consumption from D's deliberative field. Instead, C shapes D's deliberation by modifying D's reasons for drinking. Hence, D ends up choosing not to drink as a consequence of a process of rational deliberation that has been distorted by C's exaggerating drinking's dangers. And unlike when C intercedes in D's power of recognition, here C trusts D enough to allow alcohol consumption to be among the minimally choiceworthy ends D recognizes. However, C judges that D is not as competent as she is in the appraisal of alcohol's merits. C thus allows D's rational deliberation to take drinking as one of D's *potential* ends. In this respect, C's interceding in D's power of discrimination is riskier (from her point of view) than her interceding in D's power of recognition. But should C's paternalistic intercession succeed, then D's deliberation terminates in his rejecting drinking as an end. Since here C puts more trust in D's competency as a rational will or agent, C's beliefs reflect a *moderate* mistrust of the competency of D's rational will or agency.

Finally, when C intercedes in D's power of satisfaction by deceiving D regarding the invitation to the watering hole, she continues to show mistrust in D's rational agency. C endorses D's end of refraining from alcohol. To this extent, C shows greater respect for D's competency in exercising his rational will than she does when she intercedes in his powers of recognition or satisfaction. In C's estimation, D can be trusted *to deliberate*, on the basis of a factually accurate understanding of the merits of alcohol consumption, about whether to drink alcohol. All the same, C judges that D is not likely to exercise his power of satisfaction adequately. Thus C does not trust D *to act* on those deliberations, since C believes D's power to realize his chosen ends is faulty. C's beliefs thus reflect a *modest* mistrust of D's rational will or agency.

The forms of paternalism are thus as diverse as the species of rationality needed to function as rational agents. More exactly, paternalistic intercessions in the power of recognition are more morally objectionable than paternalistic intercessions in the power of discrimination because the former show greater mistrust in the targeted agent's rational competency than do the latter, and paternalistic intercessions in the power of discrimination are more morally objectionable than paternalistic intercessions in the power of satisfaction because the former show greater mistrust in the targeted agent's rational competency than do the latter. Note that I have ascribed the wrongfulness of paternalism to the level of mistrust a paternalistic act *shows*, not to the level of mistrust a paternalistic actor *feels*. For the mistrust shown by a paternalistic act need not exactly reflect the mistrust the paternalistic actor has toward her target. As stated earlier, paternalists are likely to select their forms of intercession on instrumentalist grounds, interceding in whichever power will best enable them to achieve their beneficent aims. Thus, even when a paternalist mistrusts her target's more fundamental rational power, she may nevertheless find it reasonable to intercede in a less fundamental power. In fact, on the view I have offered here, a paternalist is, all other things equal, obligated to intercede in a lesser power even when her mistrust extends to a more fundamental one. Such a paternalistic act is more justified precisely because it does *not* express the full measure of distrust she has toward her target.

Consequently, because the rational powers one mistrusts need not coincide with the rational powers one intercedes in, a paternalistic act may not fully reflect the measure of the paternalist's mistrust in her target and the target's rational powers. We must therefore be careful not

to assimilate too readily a paternalist's mistrust of her target's rational powers to the mistrust shown by her paternalistic act. The two avenues by which the tripartite structure of rational agency can explain paternalism's wrongfulness, one appealing to how these powers contribute to our identities as distinct rational agents, the other appealing to the levels of mistrust manifest in interceding in the different powers constituting that agency, are nevertheless closely linked. Judgments of trustworthiness are characterological. Paternalism involves the judgment that with respect to some matter with which another agent is concerned, that agent cannot be fully trusted. The paternalist's judgment that paternalism is warranted thus embeds a judgment about the agent targeted for paternalistic intercession. What I entrust you with is inseparable from my beliefs about who you are.

## 6. Trust, respect, and deference

A fundamental objection to my rational powers account is that the attitudes that motivate paternalistic intercessions can be epistemically warranted. Suppose that I am correct that paternalistic acts are wrong to the extent that they show mistrust in their targets' rational powers. Often though, we have evidence that others' rational powers *are* faulty—that it would be unwise for anyone, including those we might subject to paternalism, to place their trust in those powers. Self-described advocates of paternalism often appeal to this very fallibility in our powers of rational choice (Conly 2013). Why, they might ask, should we refrain from interceding in others' rational powers, given that the wrongfulness of so doing rests on its mistrust of those powers, if in fact those powers are not trustworthy? Opponents of my rational powers account may conclude that it thus requires a kind of dishonesty or make believe.

This objection can also be cast in terms of how paternalistic acts colonize or overtake our rational powers: To the extent that our identities as individuals are tied to our capacities as rational agents, we have reasons to object to being treated paternalistically because the paternalist supplants our rational agency with hers. However, it does not seem as objectionable for the paternalist to supplant our rational agency with hers if our rational agency malfunctions to such a degree that our ensuing choices and actions are not well-grounded. If our rational agency is of little instrumental value to us, its being supplanted by another's hardly counts as a loss to us.

Answering this objection requires an account of (1) exactly how refraining from paternalism values rational agency properly, and (2) why the failure to value such agency is a wrongmaking feature of paternalism even when would-be paternalists have reason to judge that their targets' rational agency is not fully trustworthy.

I grant that my rational powers account rests on an *ideal* of rational agency in two senses. First, we are not *entirely* rational. Our nature as human individuals contains other elements besides our rational agency, and not everything about us that is worthy of concern is found within that rational agency. Hence, the rational powers account idealizes inasmuch as it abstracts away from these other elements in order to highlight the special role that rational agency has in explaining the moral objectionability of paternalism. Second, we are not *necessarily* rational. Not all of our choices or decisions are made on rational grounds, and some of those that are made on such grounds are not made on good rational grounds. Rational agency is thus an ideal that we value highly despite its being only imperfectly represented in our choices and actions.

But in what regard does refraining from paternalism properly value others' rational agency? 'Value' is a verb as well as a noun, an attitude directed at objects for which that attitude is apt. In the case of rational agency, its proper valuation seems to consist in honouring boundaries between agents (even boundaries rendered more porous or pliable by particular kinds of relations among agents, such as parents and children, spouses, etc., or by professional role obligations). Paternalism, as we have seen, elides such boundaries, as one agent substitutes her judgment for another's, thereby supplanting the latter's rational powers with her own. To honour such boundaries is to show *respect* for what resides on the other side of the relevant boundary from oneself (Darwall 2006, p. 268). And as Kant saw, respect for others is an attitude marking moral distance or distinctness. Respect asks us 'not to encroach upon what belongs to anyone' (Kant 1797, p. 569), including, presumably, others' rational agency and the ends engaged by that agency.

Each of us has ends for which we are concerned. Others may not share our ends, of course. They may think, without inconsistency, that our ends rest on reasons that do not strike them as forcefully as they strike us, so that what is good enough reason for us to adopt a particular end may not be good enough reason for them to do the same. Moreover, others may refrain from meddling in our ends and our pursuit thereof for reasons unrelated to respect for us as fellow

agents. They may, for instance, simply be indifferent to what ends we pursue. Or they may fear the consequences of such meddling. Yet in expecting their respect, we are, in part, asking others to adopt a particular stance on those ends, namely, to *suppose* that those ends rest on adequate reasons. So in asking others to respect our ends, we are asking them to adopt an egalitarian stance akin to Kant's Kingdom of Ends: to see our ends and our pursuit thereof as meriting respect because they are manifestations of our rational agency, a respect that others do not show when they intercede in that agency in order either to better realize our ends or to realize ends they believe are better for us to pursue.

But in showing respect toward our cares and concerns as manifestations of our rational agency, others show respect toward that agency. If our ends reflect adequate reasons, then the powers that constitute our rational agency—the powers that enable practical deliberation—are likely to be operating as they should. That the reasons for our ends are adequate typically means that we are competent rational agents. Of course, that the reasons underlying our ends are adequate is consistent with our sometimes erring in our practical deliberations, in which case the presumed adequacy of our reasons is coincidental. Yet even then, respect for our rational agency militates against paternalism. Our rational powers are normative, rather than purely causal, powers, and so it is a constitutive fact about them that they can, at least in principle, be exercised inadequately. As we have observed, 'respect' carries a connotation of treating something as distinct and independent, as having boundaries that are not to be violated. Respect thus asks us to see the value of a thing as residing in its integrity or wholeness rather than its individual manifestations. As Isaiah Berlin pointed out, it would be a 'monstrous impersonation' of respect for rational agency to suppose that it merits respect only when it operates impeccably (2002, p. 180).

In refraining from paternalism, we treat another's will as authoritative for no other reason than that it is her will.<sup>13</sup> Hence, showing *deference* toward rational agents is an essential element of respecting rational agency, of seeing rational agency as establishing boundaries we have strong moral reasons not to trespass against. Such deference generates a high bar for the evidence needed to persuade us that someone's rational agency is faulty and thus helps to explain the

<sup>13</sup> As Groll (2014, p. 702) puts it, when a person is competent, we are to treat her will as 'structurally decisive', that is, as a source of *de jure* reasons that supersedes considerations such as her welfare, etc.



general moral presumption against paternalistic intercessions in others' agency. Notice that this does not mean that whether the target of potential paternalism is trustworthy in their practical deliberations is irrelevant to paternalism or its justification. Ian Carter has recently argued that respect:

is a substantive moral attitude that involves abstaining from looking behind the exteriors people present to us as moral agents. More precisely, while we may see behind these exteriors (for to do so is often unavoidable), if and when we do perceive people's varying agential capacities we refuse to let such perceptions count as among the reasons motivating our treatment of those people. In other words, we avoid evaluating people's agential capacities as an aid to deliberation about alternative courses of action. (2011, p. 551)

But such abstinence has a limit. At some point, evidence of another's lack of trustworthiness can raise questions about whether they are even minimally competent rational agents, in which case we descend from ideal into non-ideal theorizing about paternalism. Carter again:

Treating persons as wholly opaque, in the sense of completely ignoring their agential capacities, would be too strong, for it would preclude those assessments that are necessary in order to have the reasonable belief that they have any agential capacities at all. We perceive individuals as moral agents because we perceive them as having at least a certain minimum of agential capacities. (2011, p. 552)

The moral demand Carter dubs 'opacity respect' thus asks us to rescind from judgments about others' rational capacities so long as their capacities exceed a minimal threshold. When they do not reach this threshold, respecting rational agency gives way to valuing rational agency, i.e., seeking to promote or restore it through means that might otherwise be objectionably paternalistic (Christman 2014, p. 373).

Respect for others' rational agency is thus a moral demand that is imperfectly sensitive to our evidence concerning others' rational agency. It asks us to discount or bracket evidence we may have concerning the competence of others' rational agency in order to respect them as rational agents. It may seem odd that a moral demand can have implications regarding our epistemic obligations—that a moral stance we take toward others can imply an epistemic stance we must take toward others. But there are instances in which our moral obligations require us to *forego* the pursuit of particular knowledge or information, such as when the right to privacy entails an obligation

not to pursue certain evidence regarding another's behaviour or when the right to provide informed consent to medical interventions constrains what evidence physicians may gather in their research. The deference involved in respecting rational agency generates obligations of a different epistemic kind, compelling us (at least to some degree) to *disregard* evidence concerning an agent's rational competence. It thus operates somewhat like legal obligations, imposed on judges, to set aside evidence resting on hearsay or acquired via coercion.

In conflicts between moral normativity and epistemic normativity, the former often wins out (Wrenn 2004). In the case of paternalism, the epistemic imperfections of our rational agency do not, as some enthusiasts for paternalism would have it (Conly 2013, p. 40), mean that rational agency cannot be respected in general or when exercised inadequately. The possibility of error is built into any normative power, so that respect for rational agency requires respect for a fallible power. So our rational agency, and the choices and actions that flow from it, can and ought to be respected even on those occasions when our exercises thereof are less than sterling.

Of course, precise judgments regarding whether an individual is essentially rational but exercising her rational powers badly, as opposed to not satisfying the conditions of rational agency at all, may be difficult to come by (Cholbi 2011). A modification to my earlier example serves to illustrate this point: As far as paternalism goes, C faces a rather different set of moral challenges if, instead of trying to prevent D from picking up a drinking habit in the first place, C is attempting to stop D from drinking when D is already an alcoholic or 'problem drinker'. Should C be concerned about paternalism in the way I have represented it—as an intercession in D's rational powers—or should C conceptualize her interventions as non-paternalistic, on the grounds that while she aims to improve D's well-being, D's addiction has obliterated the rational powers intercession in which would count as paternalistic in the first place? 'Paternalistic' interventions in the choices and actions of the mentally ill raise similar worries.

Obviously, much will depend on the details of particular cases like these: how compromised the rational powers are, which powers are compromised, etc. However, my taxonomy of the rational powers can still illuminate these matters. To the extent that addicts or the mentally ill retain their rational powers, interceding in recognition is more objectionable than discrimination, which in turn is more objectionable than interceding in satisfaction, etc. And even if it is determined that the target lacks these rational powers and is therefore not a

rational agent, a moral justification for interceding is still needed despite these intercessions not rightfully being classified as paternalistic (Feinberg 1986, pp. 12-3). That an intercession in others' actions or choices is not paternalistic does not make it morally unproblematic. Such a justification may rest on the value of rational agency, but not on respecting it. The justification may instead rest on aims such as *protecting* or *promoting* rational agency. My aim here has been to show how the wrongfulness of various paternalistic acts can be mapped onto the architecture of rational agency, not to appraise possible justifications for intervening in the choices or actions of beings who are not rational agents. That 'non-ideal' project must be addressed elsewhere (Jaworska 1999; Schapiro 1999; Cholbi 2002).

## 7. The varieties of paternalism

Finally, let us note that the rational powers account helps make sense of some widely held views about the moral objectionability of different varieties of paternalism.

For instance, strong paternalism faces a higher burden of proof than does weak paternalism. Strong paternalism is directed toward a person whose choice of ends is, in the estimation of the paternalistic intervenor, mistaken or irrational. Weak paternalism, in contrast, is directed toward a person whose choices and actions will, in the estimation of the paternalistic intervenor, frustrate or undermine that person's ends. In general, strong paternalism seems more morally problematic than weak paternalism for reasons that my rational powers account illuminates. The strong paternalist mistrusts her target's capacity to select worthwhile or choiceworthy ends. Hence, the strong paternalist is likely to be more willing to intercede in those powers—the powers of recognition and discrimination—intercession in which shapes the ends a person chooses. Again, this willingness need not translate directly into interceding in these more fundamental powers. But the strong paternalist, given her willingness to shape both her target's ends and the chosen means to those ends, would presumably not hesitate to intercede in the powers associated with identifying and choosing ends (recognition and discrimination).

Weak paternalism, on the other hand, is motivated by concern not for the 'wisdom, prudence, or dangerousness' *per se* of its target's chosen ends but by concern that the target's choices and actions are, as means to her chosen ends, at odds with those ends (Feinberg

1986, p. 12). The weak paternalist thus aims to make her target's choice of action align with the target's ends. The weak paternalist is thus likely not to try to modify her target's ends but to intercede in the target's power of satisfaction so that the target's choices in fact satisfy her ends. Hence, while not all instances of weak paternalism will be less morally objectionable, either *qua* paternalistic or from an all-things-considered perspective, than all instances of strong paternalism, the view I have developed thus far explains why this will usually be the case: Strong paternalism intercedes in rational powers the intercession in which shows greater mistrust in its target's rational competency than similar weak paternalist intercessions in the power of satisfaction.

Likewise, my version of the rational will view sheds light on many philosophers' relative enthusiasm for soft over hard paternalism. Soft paternalists maintain that while we may not act paternalistically to prevent a person from knowingly and voluntarily engaging in self-harm, we may act paternalistically if doing so is necessary to determine whether a person is knowingly and voluntarily intending to self-harm. To use a timeworn example, soft paternalists do not deny that someone has the right to walk across a damaged bridge, but they permit others to interfere with her intention to do so if they have reason to believe that she is not aware of its fragile condition. Hard paternalists reject this condition, allowing paternalism to prevent self-harm even when a person knowingly and voluntarily intends such harm to herself. Hard paternalists thus assert that we may sometimes prevent others from acting on ends they have rationally chosen. Hard paternalism is thus more morally objectionable than soft paternalism, according to my version of the rational will view, because it intercedes in a more fundamental rational power—the power of discrimination—than does soft paternalism, which is aimed at ensuring that agents are exercising their power of satisfaction in accordance with their rationally chosen ends.

## 8. Conclusion

Gerald Dworkin has observed that debates about paternalism, while 'about facts ... and how effective various policies are', are also 'crucially about different ideals of the person' (2013b). Here I have argued for a version of the rational will view of paternalism—the rational powers account—that takes very seriously the ideal of the person as a rational agent, with particular attention to the architecture of

rational agency and the ways in which various paternalistic acts or policies can intercede in that architecture.

I have not taken up some disputes regarding paternalism. For example, I have not weighed in on whether paternalism intercedes in another's rational will so as to advance her interests or well-being or whether paternalism can also occur when one intercedes in another's rational will for the sake of her ends, including her non-self-interested or other-regarding ends. Nevertheless, by explaining a wide range of ethical data about paternalism in a parsimonious way, my rational powers account of paternalism appears to offer the foundation of a plausible philosophical theory of paternalism. It provides powerful explanations of why rational suasion usually deflects charges of paternalism; why consenting to intercessions in one's rational agency negates paternalism; why we ordinarily believe that strong paternalism is more objectionable than weak paternalism; and why we ordinarily believe that hard paternalism is more objectionable than soft paternalism. And although my defence of the rational powers account does not strive to provide a complete casuistry of paternalistic acts and policies, it suggests how such an account might be developed so as to make sense of the intuitively plausible claim that paternalism is presumptively, but not necessarily, impermissible. Admittedly, the rational will view's characterization of paternalism is not normatively *neutral*, since it refers to wrong or badmaking features of paternalism (Bullock 2015). However, it is also not objectionably *partial*, since it also refers to apparently goodmaking features of paternalism, namely, that (successful) paternalism helps its targets secure their ends or the ends they ought to seek. Hence, the rational will view does not illicitly close off debates about the pros and cons of particular instances of paternalism. After all, proponents of paternalism can grant that such normatively 'loaded' definitions are correct while attempting to show that the bad or wrongmaking feature is either merely *prima facie* or is outweighed by its goodmaking feature.<sup>14</sup>

## References

Anderson, Elizabeth 1999: 'What is the Point of Equality?' *Ethics* 109, pp. 287–337.

<sup>14</sup> I gratefully acknowledge the feedback and insights that David Adams, Emma Bullock, Nico Cornell, Carl Cranor, John Davis, Johannes Drerup, Megs Gendreau, Kalle Grill, Andres Moles, Steven Munzer, Chris Naticchia, Maura Priest, Jonathan Quong, Peter Ross, Peter Schaber, Danny Scoccia, Tom Walker, and three referees for this journal provided on earlier versions of this manuscript.

- Berlin, Isaiah 2002: *Four Essays on Liberty*. Oxford: Oxford University Press.
- Bullock, Emma C. 2015: 'A Normatively Neutral Definition of Paternalism'. *Philosophical Quarterly* 5, pp. 1–21.
- Carter, Ian 2011: 'Respect and the Basis of Equality'. *Ethics* 121, pp. 538–71.
- Cholbi, Michael 2002: 'Suicide Intervention and Non-Ideal Kantian Theory'. *Journal of Applied Philosophy* 19, pp. 245–59.
- 2011: 'Kantian Paternalism and Suicide Intervention'. In Coons and Weber 2013, pp. 115–33.
- Christman, John 2014: 'Relational Autonomy and the Social Dynamics of Paternalism'. *Ethical Theory and Moral Practice* 17, pp. 369–82.
- Conly, Sarah 2013: *Against Autonomy*. Cambridge: Cambridge University Press.
- Coons, C., and M.E. Weber, (eds.) 2013: *Paternalism: Theory and Practice*. Cambridge: Cambridge University Press.
- Cornell, Nicolas 2015: 'A Third Theory of Paternalism'. *Michigan Law Review* 113, pp. 1–44.
- Darwall, Stephen 2006: 'The Value of Autonomy and Autonomy of the Will'. *Ethics* 116, pp. 263–84.
- De Marneffe, Peter 2006: 'Avoiding Paternalism'. *Philosophy and Public Affairs* 34, pp. 68–94.
- Dworkin, Gerald 2013a: 'Defining Paternalism'. In Coons and Weber 2013, pp. 25–38.
- 2013b: Review of S. Conly, *Against Autonomy*. Notre Dame Philosophical Reviews (<http://ndpr.nd.edu/news/40382-against-autonomy-justifying-coercive-paternalism/>), 4 June 2013.
- Egonsson, D., J. Josefsson, B. Petersson, T. Ronnow-Rasmussen, and I. Persson, (eds) 2001: *Exploring Practical Philosophy: Essays in Honour of Ingmar Persson*. London: Ashgate.
- Feinberg, Joel 1986: *Harm to Self*. Oxford: Oxford University Press.
- Gregor, M. (ed and trans) 1996: *Practical Philosophy (Cambridge Edition of the Works of Immanuel Kant)*. Cambridge: Cambridge University Press.
- Grill, Kalle 2007: 'The Normative Core of Paternalism'. *Res Publica* 13, pp. 441–58.
- Groll, Daniel 2012: 'Paternalism, Respect, and the Will'. *Ethics* 122, pp. 692–720.

- Jaworksa, Agnieszka 1999: 'Respecting the Margins of Agency: Alzheimer's Patients and the Capacity to Value'. *Philosophy and Public Affairs* 28, pp. 105–138.
- Kant, Immanuel 1797: *Metaphysics of Morals*. In Gregor 1996, pp. 363–603.
- Kleinig, John 1983: *Paternalism*. Totowa, New Jersey: Rowman and Allenheld.
- Korsgaard, Christine 1996: *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Mill, J.S. 1859: *On Liberty*.
- Parfit, Derek 2001: 'Rationality and Reasons'. In Egonsson, Josefsson, Petersson, Ronnow-Rasmussen, and Persson 2002, pp. 17–39.
- Quinn, Warren 1994: 'Putting Rationality in Its Place'. In his *Morality and Action* (Cambridge: Cambridge University Press), pp. 228–51.
- Quong, Jonathan 2011: *Liberalism Without Perfection*. Oxford: Oxford University Press.
- Rawls, John 2005: *Political Liberalism*, expanded second edition. New York: Columbia University Press.
- Schapiro, Tamar 1999: 'What is a Child?' *Ethics* 109, pp. 715–738.
- Shiffrin, Seana 2000: 'Paternalism, Unconscionability Doctrine, and Accommodation'. *Philosophy and Public Affairs* 29, pp. 205–50.
- Tsai, George 2014: 'Rational Persuasion as Paternalism'. *Philosophy and Public Affairs* 42, pp. 78–112.
- Wrenn, Chase 2004: 'Hypothetical and Categorical Epistemic Normativity'. *Southern Journal of Philosophy* 42, pp. 273–90.