

*Note: this is the penultimate draft; please cite from the published version.

Internal and External Paternalism

Nir Ben-Moshe

University of Illinois at Urbana-Champaign (nir@illinois.edu)

Abstract: I introduce a new distinction between two types of paternalism, which I call ‘internal’ and ‘external’ paternalism. The distinction pertains to the question of whether the paternalized subject’s current evaluative judgments are mistaken relative to a standard of correctness that is *internal* to her evaluative point of view—which includes her ‘true’ or ‘ideal’ self—as opposed to one that is wholly *external*. I argue that this distinction has important implications for (a) the distinction between weak and strong paternalism; (b) the distinction between soft and hard paternalism; and (c) medical practice, where internal paternalism can, and probably often does, occur.

Keywords: Autonomy; liberty; medical ethics; soft and hard paternalism; weak and strong paternalism

1. Introduction

I introduce a new distinction between two types of paternalism, which I call ‘internal’ and ‘external’ paternalism. This distinction is motivated by two questions. First, which standard of correctness should a paternalist use in deciding whether someone’s evaluative judgments are mistaken (hence, in justifying paternalistic interference)? I will argue that there are two options, which I call ‘constitutive internal’ and ‘constitutive external’ paternalism. The standard in question could be one that is *internal* to the evaluative point of view of the paternalized subject in the following sense: it is *constituted* by the paternalized subject’s own evaluative judgments under suitable conditions, namely, those of her ‘true’ or ‘ideal’ self. But it could instead be one that is wholly *external* to that point of view (section 2).¹ The most obvious importance of this distinction is that internal paternalism is easier to justify than external paternalism; rather than imposing our own (or some other set of) values, preferences, and desires on the paternalized subject, we are protecting what the person *herself* (her true or ideal self) really values, prefers, and desires.² However, there seems to be a crucial difference between a scenario in which the paternalized subject *could* recognize the authority of her alleged true or ideal self and one in which she could *not* recognize this authority: in the former, but not the latter, scenario, the paternalizing agent respects the paternalized subject’s epistemic authority over the evaluative judgments that constitute her true or ideal self. Accordingly, one could ask—and this is the second question—if, and by what means, could the paternalized subject come to recognize the authority of her true or

¹ I borrow the terms ‘evaluative judgment’ and ‘evaluative point of view’ from Sharon Street and take them to have the following meanings: (a) evaluative judgments are “about what is a reason for what, about what one should or ought to do, about what is good, valuable, or worthwhile, about what is morally right or wrong” (Street 2006, 110); (b) the evaluative (or practical) point of view is “occupied by any creature who takes at least some things in the world to be good or bad, better or worse, required or optional, worthy or worthless [. . .] [and] who judges [. . .] that some things call for, demand, or provide reasons for others” (Street 2010, 366).

² While evaluative judgments include a person’s ends, values, and reasons, in so far as desires and preferences can determine ends and reasons, we can also include them among the constituents of “really Xing.”

ideal self? Again, I will argue that there are two options, which I call ‘epistemic internal’ and ‘epistemic external’ paternalism. It might be that the paternalized subject’s current evaluative judgments are mistaken relative to a standard of correctness whose authority she herself could, in principle, come to *recognize via deliberation*. But it could instead be that she could not come to recognize this authority simpliciter, or that she could not come to recognize it without undergoing some form of nonrational conversion (section 3). I will further argue that the distinction between internal and external paternalism has implications for (a) the distinction between weak and strong paternalism, that is, for the question of whether the paternalistic interference pertains to the means that the agent chooses to achieve their ends or to the ends themselves; (b) the distinction between soft and hard paternalism, that is, for the question of whether the paternalized subject is acting voluntarily and knowledgeably; and (c) medical practice, where internal paternalism can, and probably often does, occur (section 4). I conclude by briefly noting some limitations of my argument (section 5).

Before proceeding, I wish to clarify that my aim in this paper is not to offer my own novel analysis of the term ‘paternalism,’ but rather to make the case for the distinction between internal and external paternalism. Accordingly, I will take for granted many ideas that tend to be associated with an adequate analysis of paternalism and use them to make the case for this distinction. For the purposes of this paper, I will endorse a version of the following definition of paternalism, which, I believe, captures all of the key components of paternalistic action: P acts paternalistically towards Q if, by doing (or omitting to do) R, (1) P limits Q’s liberty or interferes with Q’s decision-making; (2) against Q’s will, without his consent, or contrary (or with indifference) to his values,

preferences, or desires; and (3) for Q's own good.³ I understand the first of the three conditions quite broadly, to include not only "interference with a person's liberty of action," as Dworkin (1972, 65) had argued in an early paper, but also interference with freedom of information or the deliberate dissemination of misinformation (Buchanan 1978, 372). Indeed, I take the first condition to include paternalistic manipulation quite generally.⁴ Regarding the second condition, while some have raised concerns about the "contrary to will criterion" (Groll 2012), for my purposes, the important idea in this condition is that P acts contrary (or is indifferent) to Q's current values, preferences, or desires (Dworkin 2015, 763); or, given my focus, that P acts contrary (or is indifferent) to Q's current evaluative judgments. Finally, while some have questioned the necessity of the third condition for a complete analysis of paternalism (Shiffrin 2000, 215–18), I will take for granted the idea that paternalistic action is done for the paternalized subject's own good. The good in question might simply be promoting what the paternalized subject *really* values, prefers, or desires—what their true or ideal self would value, prefer, and desire—although even in such instances other goods can be promoted, for example, the person's welfare or best interests.

2. Constitutive internal and constitutive external paternalism

Consider two examples of paternalism based on the belief that the paternalized subject's evaluative judgments are mistaken. First, consider a physician who administers blood products to a Jehovah's

³ This definition is taken nearly in its entirety from Scoccia (2018, 11). However, Scoccia's definition does not mention omissions and so precludes the possibility of paternalistic omissions. Thus, I have added the possibility of omitting to do R, which is in line with how Dworkin (2020, sec. 2) defines paternalism. I am grateful to an anonymous referee for alerting me to the importance of including the possibility of paternalistic omissions in the definition of paternalism.

⁴ Some have argued that paternalism cannot be defined in terms of interference with a person's liberty, since there are instances of paternalistic action that do not involve a limitation on anyone's liberty (Quong 2011, 74–76). The definition that I endorse includes the disjunctive clause "P limits Q's liberty or interferes with Q's decision-making." So, it does not assume that all instances of paternalism are necessarily defined by interference with a person's liberty.

Witness patient during an operation, despite the fact that the patient refused such products on religious grounds, because the patient's evaluative judgments are mistaken relative to their society's attitudes (according to which religious beliefs are generally mistaken). If we further assume that the physician is (also) motivated by furthering the patient's own good, this is presumably a clearcut case of paternalistic action. But now consider the following clinical case:

Mr. R., a forty-two-year-old man, who is married and has three children, was recently thrown from a horse during an equestrian competition. He landed headfirst, shattering his first and second vertebrae. This cervical spinal injury, which paralyzed him from the neck down, also halted his breathing. Mr. R. will need a wheelchair and a portable ventilator to breathe for the rest of his life. A few days after the accident, after his diagnosis and prognosis were explained to him by his doctors, Mr. R., who does not suffer from cognitive impairments and has full decisional capacity, told his wife "we should let me go," adding that he requests that his ventilator be disconnected. However, two years before the accident, Mr. R., who had entertained the possibility of a severe horse-riding accident that might leave him entirely paralyzed, had told his wife that she and their children were the most important thing to him and that he would *never* prefer death over life, no matter what his physical condition might be.⁵

Let us assume that an intern who was involved in the case thought that Mr. R.'s decision should not be respected, because Mr. R.'s identity is distorted by the trauma of being paralyzed; that is, he is not himself, in the sense that he is not thinking in terms of the values that he has long lived by.⁶ Let us also assume that the intern is advocating an intervention that limits Mr. R.'s liberty by not granting his request to be disconnected from the ventilator, a request made by an individual who has full decisional capacity. If we further assume that the intern recommends the intervention because she is (also) motivated by furthering Mr. R.'s own good, we again have a case of paternalism in which the paternalized subject's evaluative judgment is mistaken and the

⁵ While some details in this case description are fictional, others are based on Christopher Reeve's case and are borrowed from a *Wikipedia* entry that can be found here: https://en.wikipedia.org/wiki/Christopher_Reeve. Note that I am assuming that, unlike Reeve, Mr. R. is formally requesting to be disconnected from the ventilator.

⁶ I borrow this language from Brudney and Lantos (2011, 219), who use it in connection with a different case.

paternalizing agent believes that that it is legitimate to interfere in order to prevent him from acting on that judgment. However, note that in this case Mr. R.'s evaluative judgment is mistaken relative to what one might call his 'true self', to who he truly is. Indeed, the intern could also have made the case that Mr. R.'s judgment is mistaken relative to his 'ideal self', that is, to the person he could become. Thus, while in both the Jehovah's Witness and the Mr. R. examples, the ends and values in question come from "outside" the patient, there is an important difference: in the former case, they are "outside" simpliciter; in the latter case, they are merely "outside" the patient's current self.

One might act paternalistically when people make mistaken evaluative judgments, but the contrast between the Jehovah's Witness patient and Mr. R. suggests that we should ask the following question: What is the standard of correctness that makes these evaluative judgments mistaken? A person's current evaluative judgment A might fall short of a standard that is *internal* to her evaluative point of view in the following sense: A is incorrect relative to a standard of correctness that is constituted by the person's own evaluative judgments. However, and crucially, the evaluative point of view I have in mind is broader than the person's current evaluative judgments: it includes, as was the case with Mr. R., the evaluative judgments of the person's true or ideal self—of what they truly are and of what they could become—with respect to which the evaluative judgments of their current self are falling short. We could thus intelligibly ask whether we ought to act paternalistically towards Q if Q acts contrary to what she *really* values, prefers, or desires, where this is understood not in terms of the evaluative judgments of Q's current self, but rather in terms of those of Q's *authentic* or *idealized* self. Since the standard of correctness is constituted by the agent's evaluative judgments under suitable conditions, those of the true or ideal self, I will call this form of paternalism 'constitutive internal paternalism':

Constitutive Internal Paternalism: P performs an act of constitutive internal paternalism towards Q if, by doing (or omitting to do) R, P limits Q's liberty or interferes with Q's decision-making, contrary (or with indifference) to Q's current evaluative judgments and for Q's own good, because Q's current evaluative judgments are mistaken relative to a standard of correctness that is constituted by Q's own evaluative judgments under suitable conditions.

Alternatively, a person's current evaluative judgment A might fall short of a standard that is *external* to her own evaluative point of view: A is incorrect relative to a standard of correctness that is *not* constituted by the person's other evaluative judgments, even under the relevant suitable conditions. Thus, the paternalizing agent might use the values of society, as was the case in the Jehovah's Witness example, or of a broader evaluative system (or even of her own subjective values).⁷ Accordingly, I will call this form of paternalism 'constitutive external paternalism':

Constitutive External Paternalism: P performs an act of constitutive external paternalism towards Q if, by doing (or omitting to do) R, P limits Q's liberty or interferes with Q's decision-making, contrary (or with indifference) to Q's current evaluative judgments and for Q's own good, because Q's current evaluative judgments are mistaken relative to a standard of correctness that is *not* constituted by Q's own evaluative judgments under suitable conditions.⁸

While I will not be offering a systematic defense of the notion of a true or ideal self, both notions, which pertain to what an agent *really* values, prefers, or desires, are familiar ones. The notion of a true self involves, at the very least, the following three claims: (a) people tend to have a set of somewhat stable and coherent core evaluative judgments; (b) these core evaluative judgments are most deeply the agent's own, which is a function of how strongly she holds them

⁷ One could play with the details of the Jehovah's Witness example so that the standard in question is derived from a broader evaluative system. This would be the case if, for example, a Kantian or utilitarian evaluative system would show that, all things considered, the patient's evaluative judgments are mistaken.

⁸ It is helpful to think of the distinction between internal and external paternalism by reference to a version of one of Shiffrin's (2000, 218) conditions on paternalistic behavior—namely, that such behavior involves the substitution of P's judgment or agency for Q's. While in cases of constitutive external paternalism, P substitutes her own (or others') evaluative judgments for Q's evaluative judgments, in cases of constitutive internal paternalism, P substitutes the evaluative judgments that Q *herself* would make under suitable conditions for Q's current evaluative judgments.

and how close to the center of her total web of evaluative judgments they lie; (c) living in accordance with the set of these core evaluative judgments is an important good for the person.⁹ Indeed, sometimes life might not be worth living if it cannot be lived according to these evaluative judgments. This is so because, among other things, the idea of a true self is closely associated with the ideal of *authenticity*, namely, the capacity to be a particular self (or “oneself”) and live a distinctive life in accordance with one’s core evaluative judgments (Brudney 2009, 32; Brudney and Lantos 2011, 221). If a person is prevented from living an authentic life, so the thought continues, she will lose her individuality and integrity (e.g., Dworkin 1993; MacIntyre 1981; Taylor 1991). The notion of an ideal self, a self which is also under certain suitable conditions, is familiar from dispositional theories of values and Humean theories of reasons: for example, something is of value if and only if one is under suitable conditions that include full imaginative acquaintance with that thing (Lewis 1989); or one’s evaluative judgment that p is a reason for one to ϕ is correct if and only if that judgment is capable of withstanding scrutiny from the standpoint of one’s other evaluative judgments in reflective equilibrium (Street 2006, 2008). Therefore, one can understand the self that has, for example, full imaginative capacity or whose evaluative judgments are most coherent in reflective equilibrium as one’s ideal self. Importantly, the introduction of the pertinent suitable conditions leaves room for evaluative error, for an agent can value something independently of whether she presently values it. Nevertheless, the standard of

⁹ I borrow claims (a) & (c), with some alterations, from Brudney & Lantos (2011, 224). A version of claim (b) can be found in Street (2008, 235). I have tried to keep the metaphysical assumptions pertaining to the true self to a bare minimum: I am not advancing the idea of a metaphysically-suspect “essence” that is the agent’s true self. Nevertheless, given the combination of claims (b) and (c) above, I understand the evaluative commitments of one’s true self to be both genuine *and* significant: if these two dimensions were to come apart—if one were to allow, for example, the true self to encompass genuine, but not especially significant, evaluative commitments—the very idea of a true self would be trivialized. See also Enoch (2017, 23-28) for a discussion of the role that one’s true self—or ‘depth,’ ‘alienation,’ ‘centrality to the self,’ and ‘higher-order desires’—may play in connection with hypothetical consent.

correctness that determines what counts as evaluative error is ultimately set by the agent herself when she is under the relevant suitable conditions (Street 2006, 110–11; 2008, 207–8).

Three things are worth noting. First, one's true and ideal selves could coincide: one's deepest evaluative judgments might be those of one's ideal self and hence one might act most authentically in the role of one's ideal self. But the two selves are conceptually distinct and can come apart: the evaluative judgments that are most deeply one's own need not be those that one has under conditions of, for example, imaginative acquaintance or coherence. Second, insofar as one might worry that the idealized standard involved in the idea of an ideal self is not necessarily the agent's own, the following observation may be of help: the idealization in question is very modest in the sense that the agent's ideal self simply *is* the current self with greater imaginative acquaintance or greater coherence. Thus, I am also *not* assuming that we all have the same idealized self; rather, each agent can have different values, preferences, and desires, even under suitable conditions. Third, the type of evidence we would need in order to make judgments about an agent's true self is, in most circumstances, *backward-looking*: the agent's current evaluative judgments are mistaken relative to her *previous* evaluative judgments, where the latter are indicative of her true self. By contrast, the type of evidence we would need in order to make judgments about an agent's ideal self is, in most circumstances, *forward-looking*: the agent's current evaluative judgments are mistaken relative to her *future* evaluative judgments, where the latter are indicative of her ideal self. In the latter type of case, the type of evidence we might require includes the fact that the paternalized subject has yet to use their imagination or that there are inconsistencies in their current evaluative judgments that would give one reason to conjecture that their ideal self may be more coherent. A key assumption pertains to the value of epistemic humility: the thought is that we have good reason to *suspect* that the agent may be mistaken relative to her ideal self. So, for example,

Mr. R. might insist that he would like to be disconnected from the ventilator as soon as possible. However, he might also insist that he most values, for example, seeing his son get married next year, or seeing his daughter give birth next month, or completing a book of fiction he is writing. We may not know what Mr. R.'s true self is in regard to these issues, but we can hope that his ideal self will be more coherent. Needless to say, we cannot know for certain that Mr. R.'s future evaluative judgments will be more coherent, much in the same way that we cannot know for certain that his past evaluative judgments are indicative of his true self; but these are just some of the *epistemic* complexities associated with trying to identify a person's true or ideal self.

3. Epistemic internal and epistemic external paternalism

The most obvious importance of the internal/external distinction is that internal paternalism is easier to justify than external paternalism, since the limitations on or interferences with the paternalized subject are less substantial in the former than they are in the latter. Indeed, rather than imposing our own (or some other set of) values, preferences, and desires on the paternalized subject, we are protecting what the person *herself* (her true or ideal self) really values, prefers, and desires. The thought is that if the paternalized subject had correctly deliberated from the evaluative perspective of her own true or ideal self, she would have decided to do precisely what the paternalistic intervention forced to her to do. Therefore, from the perspective of her true or ideal self, the intervention simply saved her from making a mistake—a mistake she would have made only because of the gap between her true or ideal self and her current self. Now, it might be objected at this point that the fact that some truer or more idealized version of yourself would prefer some other end or value than what you presently prefer may not strike you as relevant. For example, in medical contexts, one may well resent being subjected to physical assault (unwanted

treatment, for instance) on the basis that this is what one's true or ideal self would prefer. In other words, why should an agent be any less opposed to being subjected to the values of their true or ideal selves than being subjected to the values of someone else altogether?¹⁰ Here there seems to be a crucial difference between a scenario in which the paternalized subject *could* recognize the authority of her alleged true or ideal self and one in which she could *not* recognize this authority. Surely, if the paternalized subject could recognize the authority of her true/ideal self, then the paternalism in question would be easier to justify; after all, in such cases, the paternalizing agent respects the paternalized subject's epistemic authority over the evaluative judgments that constitute her true or ideal self. In other words, if the paternalized subject could *herself* recognize the authority of the standard in question—via certain means, as will become apparent shortly—then, not only is it the case that the verdicts of the true or ideal self are authoritative because this self is under the relevant suitable conditions (and, as I will suggest in section 4, is also potentially more autonomous)—although they *are* also authoritative for this reason—but, in addition, they are authoritative because the true or ideal self would endorse the deferral to itself as the authoritative standard. This self has, therefore, the requisite (and conclusive) authority to approve of the paternalistic intervention.¹¹

¹⁰ While its focus is usually on patients who are currently incompetent, for example, due to dementia, similar worries are familiar from the literature on precedent autonomy. That literature has considered, for instance, whether the provisions of a prior advance directive should be honored, despite a patient's current opposition to their being honored, because the prior self is more authentic or authoritative. While Dworkin (1993, chap. 8) famously defends deferring to the patient's past self since it is more authentic and authoritative, Dresser (1995), for example, makes a case for respecting the incompetent patient's current interests. I am grateful to an anonymous referee for posing the worry discussed above as well as for drawing my attention to this (partial) analogy to the precedent autonomy literature.

¹¹ To return to the analogy with the precedent autonomy literature, my position is (partially) analogous to Davis's (2004) suggestion, in support of a Dworkin-type position, that when a person has two or more conflicting preferences, we must respect any third preference she has about how to resolve the conflict. Since, as Davis further notes, these preferences need not exist at the same time, nor does respect for a preference have to be shown at the same time the preference exists, our duty to respect a preference formed by a competent informed agent is not undermined by the fact that the preference exists before or after the time it is respected. To be sure, Davis focuses on specific preferences and I focus on the authority of the true/ideal self. However, the reasoning is similar, especially

Of course, there are various ways in which a person might be brought to recognize the authority of her true or ideal self: she might, for example, be hit on the head with a hammer or undergo some form of nonrational conversion. Presumably it is easier to justify the paternalism in question if the paternalized subject could recognize the authority of her true or ideal self in a way that respects her rational capacities, that is, via *deliberation* from her actual self. The key thought is that, in such cases, the agent could, in principle, deliberate from the evaluative judgments of her actual self to those of her true or ideal self and endorse them.¹² So one could ask if, and by what means, the paternalized subject could come to recognize the authority of her true or ideal self, and, more generally, of the standard of correctness in question. Focusing on the more general form of the question, it is helpful to define what I will call ‘epistemic internal paternalism’ in a more general way—that is, in a manner that abstracts away from talk of a ‘true’ or ‘ideal’ self:

Epistemic Internal Paternalism: P performs an act of epistemic internal paternalism towards Q if, by doing (or omitting to do) R, P limits Q’s liberty or interferes with Q’s decision-making, contrary (or with indifference) to Q’s current evaluative judgments and for Q’s own good, because Q’s current evaluative judgments are mistaken relative to a standard of correctness whose authority could be recognized by Q following deliberation.¹³

if we note that Davis’s thought is, presumably, that (a) the *competent* self is authoritative (which can be construed as a self under relevant suitable conditions and which is thus also more autonomous), and (b) in making an all-things-considered judgment regarding specific preferences, this authoritative self would endorse the deferral to itself as the authoritative standard.

¹² I borrow this thought from Williams (1995, 35; 2001, 91), who argued that *p* is a reason for A to ϕ only if A could reach the conclusion that he should ϕ by a sound deliberative route from his motivations. It is worth noting, however, that Williams (and Sharon Street), would presumably not draw a distinction between reasons that are constituted under suitable conditions and our ability, at least in principle, to recognize these reasons following deliberation. However, this distinction is conceptually possible and, I believe, normatively significant in the context of paternalistic action.

¹³ While my focus in this section is on the ability of the paternalized subject to recognize the authority of his true or ideal self, and hence on constitutive internal paternalism, the more general formulation of epistemic internal paternalism above—and, indeed, epistemic external paternalism below—can also be applied to cases of constitutive external paternalism. So, for example, in the Jehovah’s Witness case, one might also question whether the patient could recognize the “correctness” of society’s values following deliberation (though presumably he will not do so).

Note that ‘could’ plays a crucial role in my definition of epistemic internal paternalism, for, given the current condition of the paternalized subject, she might not presently deliberate from the evaluative judgments of her current self to those of her true or ideal self. This might be the case with Mr. R., who, while having full decisional capacity, may be too preoccupied with recent developments in his life to deliberate towards the values, preferences, and desires that define who he really is or could become. Indeed, if Mr. R. were *immediately* able to deliberate from the evaluative judgments of his current self to the conclusion that he ought to accept or continue treatment, it is far from clear that there would be any need for paternalistic action. Therefore, it is *not* the case that once we grant the epistemic dimension of internal paternalism, the role of paternalism is superfluous: while it is true that once the paternalized subject recognizes the authority of the standard in question, he is consenting to whatever that true or ideal self would recommend (in which case paternalism is no longer at issue), it is nevertheless plausible that certain forms of paternalistic intervention can be utilized *prior* to the recognition of this authority.

Three points are worth noting. First, if the paternalizing agent is sincere in her belief that the paternalized subject could recognize the standard of correctness in question following deliberation, the former would presumably limit her paternalistic interventions to those that do not undermine the latter’s ability to deliberate appropriately. So, the paternalistic intervention could focus only on, for example, interfering with the paternalized subject’s ability to *execute* certain actions. This is what I proposed in regard to Mr. R.—namely, limiting Mr. R.’s liberty by simply not granting his request to be disconnected from the ventilator. In other cases, *minor* interferences with the paternalized subject’s rational powers—interferences that do not undermine the latter’s ability to recognize the authority of the standard of correctness in question following deliberation—might

be warranted.¹⁴ Second, a judgment of the form “Q *could* recognize the authority of their true or ideal self” is context-sensitive in the following sense: the point in time when the paternalizing agent should concede that the paternalized subject could *not* recognize the authority of the standard in question will depend on the latter’s psychology and on the situation more broadly. Accordingly, there will be a point at which, if the paternalizing agent continues to insist that certain evaluative judgments are reflective of the paternalized subject’s true or ideal self, this insistence may be a mistake or, perhaps, mere bluff. So if Mr. R. has processed his initial emotional reactions to his new life situation by, for example, talking through the situation over an extended period of time, but still prefers, all things considered, to die, then his physicians should concede that his stated preferences may in fact reflect what he *really* values, prefers, and desires.¹⁵ Third, if we return to the worry that the *idealized* standard involved in the idea of an ideal self is not necessarily the agent’s own, we can now add that on epistemic internal paternalism, the actual self could recognize the authority of the alleged ideal self following deliberation. Therefore, the actual self could recognize itself in the ideal self, while at the same time recognizing that the ideal self is better situated to determine what the agent *really* values, prefers, or desires. So, if one is a constitutive

¹⁴ Cholbi (2017) distinguishes between three rational powers: (a) the power of recognition (recognizing ends as minimally choice worthy); (b) the power of discrimination (rationally selecting which ends to endorse); and (c) the power of satisfaction (pursuing those means to realize our endorsed ends). If we alter (c) so as to accord with the distinction offered in the current paper, we could understand the power of satisfaction to include the ability to pursue those means to realize the ends *of the true or ideal self*. Accordingly, paternalistic intervention with this rational power need not interfere with the paternalized subject’s very ability to recognize or rationally select various ends. To use Cholbi’s examples, there need be no elimination of ends from the menu of ends from which the paternalized subject can choose, nor a tilting of the scales against actions by making them less choice worthy (for example, by exaggerating their dangers). Rather, the paternalistic intervention could interfere with the paternalized subject’s ability to pursue the ends of the actual self in limited ways: for example, by hiding opportunities that would allow him to satisfy some of the ends of the actual self. Presumably, this type of intervention, which is a means of allowing the paternalized subject to realize the ends of his true or ideal self, need not undermine the latter’s ability to deliberate appropriately.

¹⁵ Despite the context sensitivity, one can put safeguards in place such that the paternalizing agent does not abuse his power. In medical contexts, if physicians employ internal paternalism, they should be required to constantly explain to relevant others—who can hold them accountable—why they think that the paternalized subject could recognize the authority of her true or ideal self, even if she has yet to do so.

and epistemic internal paternalist, one will maintain that the paternalized subject's ideal self is under the relevant suitable conditions for determining that person's authoritative values, preferences, and desires *and* that these conditions are easily attainable for the person's actual self such that this self could recognize (via deliberation) the authority of the values, preferences, and desires of his ideal self. (And one could make an analogous case regarding the relations between the agent's actual and true selves.)

Epistemic internal paternalism respects the thought that what the agent really values, prefers, and desires—that is, what their true or ideal self would value, prefer, and desire—is something she herself could recognize following deliberation. There is, however, another possibility: the paternalizing agent does in fact recognize the evaluative judgments of the paternalized subject's true or ideal self, but the paternalized subject could not recognize the authority of her own true or ideal self even following deliberation. Perhaps there is an inability on the part of the paternalized subject to come to recognize the authority of the normative standard in question *simpliciter*. However, perhaps the paternalized subject could come to recognize the authority of the normative standard, but only following a nonrational conversion. The latter possibility is related to a well-known exchange between Bernard Williams and John McDowell about the existence of external reasons. While Williams (1981, 108–9) ruled out the possibility of an agent coming to believe an external reason statement via nonrational alterations to his motivations, McDowell (1998, 102, 107) argued that a person might come to recognize external reasons as *their* reasons only after such alterations. Abstracting away from talk of a true or ideal self, I will call the general position according to which the paternalized subject could *not* recognize the authority of the standard of correctness in question following deliberation 'epistemic external paternalism':

Epistemic External Paternalism: P performs an act of epistemic external paternalism towards Q if, by doing (or omitting to do) R, P limits Q's liberty or interferes with Q's decision-making, contrary (or with indifference) to Q's current evaluative judgments and for Q's own good, because Q's current evaluative judgments are mistaken relative to a standard of correctness whose authority could *not* be recognized by Q following deliberation.¹⁶

When focusing on Q's true or ideal self, there is an assumption that the evaluative judgments of her true or ideal self have normative force for her regardless of whether she could recognize the correctness of these judgments following deliberation. Paternalism based on this assumption should be used with caution because the paternalized subject could not recognize the ends and values in question as her own, much in the same way that she could not recognize the values of society, for example, as her own, if they were imposed on her. (One could argue, however, that the former case is a weaker form of external paternalism than the latter, since, given their origin in the agent's true or ideal self, the values in question are more "internal" to the agent than those in the latter case, where the values of society are imposed entirely from "outside" the agent's evaluative point of view.) Ultimately, the paternalism that is easiest to justify is a comprehensive form of internal paternalism, which includes both the constitutive internal *and* the epistemic internal dimensions.¹⁷

¹⁶ Again, I understand 'could not' in this context to be the result of either (a) an inability to recognize the authority of one's true or ideal self simpliciter, or (b) an inability to recognize this authority without undergoing some form of nonrational conversion.

¹⁷ The distinction between epistemic internal and external paternalism, as I am using it, focuses on the question whether, and by what means, the paternalized subject could recognize the authority of her true or ideal self; it does not focus on the question whether, and by what means, the paternalized subject comes to understand the relations between a proposed course of action and the evaluative judgments of her true or ideal self. Thus, this distinction is not intended to track paternalism in cases in which, for example, the person's true self values improved health and treatment X will, in all probability, contribute to improved health, but, since she is prone to errors in statistical reasoning, this person will never recognize the value of treatment X for her. Furthermore, the distinction is not intended to track paternalism if it were the case that, for example, the person in question will come to view the treatment as good, as in line with her true self, but that she does so only because of fallacious statistical reasoning to which she is prone, that is, for the wrong reason. I am grateful to an editor at the *Canadian Journal of Philosophy* for these points.

4. Implications

Apart from the different levels of justification involved in internal and external paternalism—internal paternalism being a weaker form of paternalism than external paternalism—this distinction has several important implications. Perhaps its most important implication is for the distinction between weak and strong paternalism. While a weak paternalist believes that “it is legitimate to interfere with the means that agents choose to achieve their ends, if those means are likely to defeat those ends,” as Dworkin (2020, sec. 2.3) puts it, a strong paternalist believes that “people may have mistaken, confused or irrational ends and it is legitimate to interfere to prevent them from achieving those ends.” The difference, as Dworkin continues, is that the weak paternalist targets “mistakes about the facts,” whereas the strong paternalist targets “mistakes about values.” The importance of the distinction lies in the fact that it is easier to justify weak paternalism than it is to justify strong paternalism, since the former, but not the latter, respects the ends and values of the paternalized subject. Accordingly, when Dworkin defines weak paternalism, he provides the following example: “if a person really prefers safety to convenience then it is legitimate to force them to wear seatbelts” (sec. 2.3). Likewise, Conly (2013, 2) speaks of helping “one another avoid mistakes so that we may all end up where we want to be,” where the paternalism in question is, presumably, weak paternalism. Indeed, the paternalism that Conly promotes is not paternalism about “objectively good ends, or objectively rational ends,” ends that everyone should be made to pursue. It is, instead, about mere means: on her view, paternalistic behavior is called for only when “people’s choices of instrumental means are confused, in a way that means they will not achieve their ultimate ends.” In other words, “interference is justified on paternalistic grounds only when it reflects individuals’ actual values, not the values we might like them to have” (Conly 2013, 43, 150). Therefore, weak paternalism is supposed to respect what a person “really prefers,”

their “ultimate ends” and “actual values,” and where they ultimately “want to be.” Strong paternalism, on the other hand, presumably imposes ends and values that are foreign to the paternalized subject, ends and values that come from “outside” the person, as it were—which is harder to justify.

Since they both pertain to interfering with people’s allegedly mistaken evaluative judgments (or ends and values), internal *and* external paternalism can be construed as forms of *strong* paternalism. And since internal paternalism is weaker than external paternalism, the internal/external distinction can enrich our understanding of strong paternalism and allow for finer-grained distinctions within this category. (Recall that talk of ends and values coming from “outside” the person can mean two quite different things: “outside” simpliciter versus “outside” the person’s current self.) However, note that the reason that weak paternalism is easier to justify than strong paternalism is of the same type as the reason that internal paternalism is easier to justify than external paternalism: both weak *and* internal paternalism respect the ends and values of the paternalized subject. Indeed, most of the key observations that Dworkin and Conly make about weak paternalism are also true of internal paternalism: after all, the intern, who is advocating internal paternalism in Mr. R.’s case, is also respecting what Mr. R “really prefers,” his “ultimate ends” and “actual (true or idealized) values,” and where he ultimately wants to be. This is surprising since, as noted, internal paternalism, like external paternalism, is, on the face of it, a form of *strong* paternalism. Thus, paternalism that protects what an agent “*really* values, prefers, or desires” may arise either when there is no question of evaluative error on the part of the paternalized subject (weak paternalism), or when there is a question of evaluative error on the part of the paternalized subject (strong paternalism of the internal type). Accordingly, paternalists interested in the weak/strong paternalism distinction should be careful in differentiating between

two questions when claiming that the paternalizing agent protects what the paternalized subject *really* values, prefers, or desires: Is the former taking the latter's current ends and values as correct and helping him take the relevant means in order to attain them? Or is the former questioning the correctness of the latter's current ends/values in light of those of the latter's true or ideal self?¹⁸ Of course, this assumes that the distinction between internal and external paternalism is a distinction *within* strong paternalism and that internal paternalism, as a form of strong paternalism, should be contrasted with weak paternalism. However, the fact that most of the key observations that Dworkin and Conly make about weak paternalism are also true of a certain kind of strong paternalism could suggest that the weak/strong paternalism distinction is not really doing substantive work. Therefore, there is a case to be made that paternalists should forget about the weak/strong paternalism distinction altogether and ask whether the paternalized subject's current evaluative judgments are mistaken relative to a standard of correctness that is internal or external to her evaluative point of view; they should then understand what has been called 'weak paternalism' as a limiting case of internal paternalism in which the current self simply *is* the true or ideal self.¹⁹

The distinction between internal and external paternalism also has implications for the distinction between soft and hard paternalism, a distinction that goes back to Mill:

¹⁸ Other authors, who do not necessarily focus on the distinction between weak and strong paternalism, also overlook these complexities: Savulescu (1995, 329) merely notes the difference between patients failing "to make choices which best satisfy their own values" and patients making "incorrect value judgments"; and Thaler and Sunstein (2009, 5) argue for the idea of making "choosers better off, *as judged by themselves*" in the context of libertarian paternalism.

¹⁹ There is reason to retain the weak/strong paternalism distinction if the distinction between the absence and presence of evaluative error is meaningful and normatively significant. Some authors, such as Groll (2019), have put pressure on the relevance and soundness of the means/ends and fact/value distinctions for the types of cases that are usually discussed in connection with weak and strong paternalism. While it is beyond the scope of this paper to examine them, if such arguments are successful, then there may be an even stronger case that the internal/external distinction should *replace* the weak/strong distinction. I am grateful to an anonymous referee for suggesting the latter possibility.

If [someone] saw a person attempting to cross a bridge which had been ascertained to be unsafe, and there were no time to warn him of his danger, they might seize him and turn him back, without any real infringement of his liberty; for liberty consists in doing what one desires, and he does not desire to fall into the river. Nevertheless, when there is not a certainty, but only a danger of mischief, no one but the person himself can judge of the sufficiency of the motive which may prompt him to incur the risk: in this case, therefore, (unless he is a child, or delirious, or in some state of excitement or absorption incompatible with the full use of the reflecting faculty), he ought, I conceive, to be only warned of the danger; not forcibly prevented from exposing himself to it. (1989, 96–97)

According to Mill, it is legitimate to limit a person’s liberty or interfere with their decision-making only if the person is deficient in knowledge or rational faculties; that is, only if it is necessary to determine whether the paternalized subject is acting “voluntarily and knowledgeably,” as Dworkin (2020, sec. 2.1) puts it, which has come to be known as ‘soft paternalism.’ According to hard paternalism, it is legitimate to limit a person’s liberty or interfere with their decision-making *even if* the person is acting voluntarily and knowledgeably.²⁰ Now, one might argue that insofar as agents are acting voluntarily and knowledgeably, all instances of internal paternalism, even though they involve less limitation of a person’s liberty and/or interference with their decision-making than those of external paternalism, are instances of hard paternalism and so are normatively suspect. A *prima facie* straightforward response to this worry is to argue that internal paternalism protects the agent from acting on misinformation in the same way that coercive action in Mill’s bridge example protects the agent from acting on misinformation: while Mill presents an example of an agent who is misinformed about certain facts, internal paternalism pertains to agents who are misinformed about certain values. However, this analogy will not do for two reasons. First, external paternalists can also argue that they are coercing the agent because she is mistaken about

²⁰ Different authors may be focusing on somewhat different issues when discussing the soft/hard paternalism distinction. As Hanna (2018) shows, while some authors focus on the notion of voluntariness, others focus on the notion of instrumental rationality. Discussing these nuances is beyond the scope of the current paper—my own focus will be on the notion of autonomy (which has, of course, relations to both voluntariness and rationality).

values; these values just happen to be external to the agent's evaluative point of view. Second, soft paternalism is often understood as applying primarily to cases of *incompetent* agents (Dworkin 1988b, 124), or, at the very least, to actions that are not perfectly voluntary because the agent is suffering from “compulsion, misinformation, excitement or impetuosity, clouded judgment [. . .], or immature or defective faculties of reasoning” (Feinberg 1971, 110–11). Mr. R., who has full decision-making capacity, is not incompetent; nor is it the case that his decisions are distorted in these ways. Therefore, it would seem that were we to limit Mr. R.'s liberty, we would be, again, hard paternalists. Nevertheless, I wish to suggest that there is a case to be made that internal paternalism corresponds to a version of soft paternalism, thereby broadening our conception of soft paternalism. The case for this correspondence relies on both the constitutive and the epistemic dimensions of internal paternalism, as well as on the thought that internal paternalism, much like soft paternalism, may promote *autonomy*: they both protect agents from choices that are not fully their own.²¹ Indeed, given the relations between the true self and authenticity, internal paternalism, which can promote the evaluative judgments of the true self, can also promote authenticity.²²

While Mill writes that “liberty consists in doing what one desires,” one could argue that we are most autonomous not when we act on the evaluative judgments that we happen to have, but when we act on evaluative judgments made *under suitable conditions*.²³ Thus, it is reasonable to assume that we are generally more autonomous when our choices are authentic—when they reflect who we truly are. So if, for example, a Jehovah's Witness suddenly agrees to accept blood products, we have *prima facie* reason to question whether the choice is fully autonomous given the

²¹ For a discussion of the idea that soft paternalism promotes autonomy, see, for example, Pope (2004, 670).

²² For a discussion of how paternalism might *undermine* authenticity, see, for example, Conly (2013, 80–85).

²³ Developing this thought is beyond the scope of the current paper. An example of how to understand “suitable conditions” in this context is Frankfurt's (1971) use of higher-order desires in order to make sense of endorsement. Indeed, Dworkin (1988c, 15–20) and Enoch (2017, 25–28) apply this Frankfurtian strategy specifically to autonomy.

uncharacteristic nature of this choice.²⁴ It is also reasonable to assume that we are more autonomous when our evaluative judgments are more ideal on certain dimensions, for example, by being more coherent. Accordingly, if a Jehovah's Witness insists that he most values conforming to God's will even at the cost of prospective death, but also says that he most values seeing his daughter graduate from college next spring, we have *prima facie* reason to question whether the choice is fully autonomous given the blatant inconsistency in his evaluative judgments. Therefore, constitutive internal paternalism can make the paternalized subject more autonomous. Of course, as is well-known, deliberation can also make agents more autonomous by making agents think through the reasons for their choices and actions.²⁵ Hence, epistemic internal paternalism can also make the paternalized subject more autonomous. This observation about deliberation is not entirely independent of the observation that making evaluative judgments under suitable conditions renders one more autonomous; as Feinberg (1971, 111) himself notes, "chosen actions are those that are decided upon by *deliberation*, and that is a process that requires time, information, a clear head, and highly developed rational faculties. [. . .] [They] not only have their origin 'in the agent,' they also represent him faithfully in some important way: they express his settled values and preferences." Accordingly, on a slightly expanded, albeit not unreasonable, conception of soft paternalism, internal paternalism is a form of soft paternalism. The conception in question includes not only incompetent agents or those suffering from compulsion, clouded judgment, defective faculties of reasoning, and so on, but also agents whose choices are not fully their own because they fall short of the choices of their true or ideal self and/or are not the result

²⁴ Self-determination and authenticity are often evidence for each other's presence: we can question a person's self-determination if her choices are substantially out of line with who she is and we can question whether a person's choices are authentic if her self-determination is jeopardized (Brudney 2009, 32). For additional discussion of the relations between authenticity and autonomy more generally, see Meyers (1989, part 2, especially sec. 2–3).

²⁵ For example, when Savulescu (1995, 330) presented his "rational noninterventional paternalism," according to which physicians should make all-things-considered value judgments and rationally argue with patients, he argued that deliberation can make patients more autonomous by making the patient better informed of both facts and values.

of deliberation. (Constitutive/epistemic) external paternalism, on the other hand, which does not promote choice in accordance with the agent's own (true or idealized) values and/or deliberation, does not make people more autonomous in any obvious sense and is, therefore, a hard form of paternalism.²⁶

Finally, while the distinction between internal and external paternalism can have implications for various intimate relationships, it has, as Mr. R.'s case has shown, important implications for the medical realm in particular. This is especially true in light of a well-known distinction in the medical ethics literature between a person's self-determination and her best interests, between an agent's capacity for agency, in the form of the ability to make choices on the basis of reasons, and her capacity to live a good life by satisfying a certain set of interests, respectively (Brudney 2009, 32; Brudney and Lantos 2011, 220). A physician might thus act paternalistically, undermining the patient's self-determination, in order to protect a patient's best interests. This usually happens when the patient is acting contrary to their *medical* interests in one or both of the following ways: (a) the patient is taking the wrong means to promoting her own health-related interests (weak paternalism); (b) the patient is wrong about her interests in the sense that health-related interests are not among them or are not ranked sufficiently high among them (strong paternalism).²⁷ However, the distinction between self-determination and best interests is insufficient to capture

²⁶ Two points are worth noting. First, some authors have argued that soft paternalism may not be "paternalistic" at all (e.g., Feinberg 1986, 12). However, I believe that internal paternalism is in fact paternalistic, since we are promoting Q's autonomy by limiting her liberty or interfering with her decision-making. For discussion of a similar point, see Dworkin (1988a, 106). Second, my claim that internal paternalism is a form of soft paternalism does not blunt the distinction between internal and external paternalism. In particular, I am not arguing that internal paternalism can be *defined* in terms of soft paternalism, nor that it is *coextensive* with soft paternalism, but rather that it can be viewed as falling *under the category* of soft paternalism (if we broaden our conception of soft paternalism in a meaningful way).

²⁷ The discussion above refers to patients who are sufficiently competent to make the decision in question. If a patient's self-determination is substantially undermined, the patient might not be sufficiently competent to make the decision. Accordingly, in such cases, the physician may also act paternalistically to protect the patient's best interests.

the type of value-related questions that may arise in clinical practice. In particular, the patient's true/ideal values, preferences, and desires—those of their true/ideal self—are not identical either to the mere ability to choose or to living a good life by satisfying certain interests.²⁸ And in terms of paternalistic action, the implication is twofold. First, while acting paternalistically for the sake of the patient's best interests might be an external form of paternalism, doing so for the sake of the patient's true or ideal self is an internal form of paternalism: while some of a person's best interests could be external to her evaluative point of view—for example, one might not value one's medical interests even under suitable conditions and/or not recognize them even after deliberation—the evaluative judgments of a person's true or ideal self should be (constitutively/epistemically) internal to it. Second, while acting paternalistically in order to protect a patient's medical interests usually requires that the physician utilize his medical knowledge—construed broadly to include both scientific/technical knowledge and knowledge of health-related values—this type of knowledge is unlikely to do much work in cases of internal paternalism. Therefore, it is important that physicians appreciate the differences between internal and external paternalism, as well as the extent to which each form of paternalism may or may not rely on medical knowledge and medical values. And insofar as physicians act paternalistically in the internal sense, they should do so if (a) they, or relevant others such as surrogates, know the patient quite well and can identify pertinent features of his psychology, primarily his authentic or idealized evaluative judgments; (b) they, or relevant others such as surrogates, engage in sustained deliberation with the patient before and/or after any paternalistic action; and (c) they are wary of imposing their own values on the patient.

²⁸ For example, focusing on the true self: while acting in accordance with one's true values, preferences, and desires furthers one's best interests, it is only one component of those interests, for such action can be at odds with one's *overall* best interests by making for a worse, albeit more authentic, life (Brudney 2009, 35).

5. Conclusion

I have introduced a new distinction between two types of paternalism: (a) acting paternalistically toward others because their current evaluative judgments are mistaken relative to a standard of correctness that is constituted by their own evaluative judgments under suitable conditions and/or whose authority they could recognize following deliberation (constitutive/epistemic internal paternalism); and (b) acting paternalistically toward others because their current evaluative judgments are mistaken relative to a standard of correctness that is *not* constituted by their own evaluative judgments under suitable conditions and/or whose authority they could *not* recognize following deliberation (constitutive/epistemic external paternalism).

It is worth noting three limitations of the main argument of this paper. First, my argument is premised on the notions of true and ideal selves, and on their normative relevance. Accordingly, readers who are skeptical of these notions may be unmoved by the distinction that I have introduced. Nevertheless, as noted, the importance of the notions of true and ideal selves have been recognized by many philosophers, and reasoning that is premised on these notions is likely to occur in various contexts in which paternalistic action can (and does) take place. Second, there are difficulties involved in identifying a person's true or ideal self. It is often far from clear who a person truly is or who she could ideally become. The paternalizing agent needs to be open-minded about, for example, the possibility that the paternalized subject might not have had core evaluative commitments pertaining to the decision at hand, or that her core evaluative commitments may have changed over time, or that she might not have a more coherent set of evaluative commitments following deliberation. The paternalizing agent should constantly ask himself whether it might not be the case that he is simply projecting his own values, preferences, and desires on to the situation and falsely claiming that they are the values, preferences, and desires of the paternalized subject's

true or ideal self. Third, I have not provided any specific justification for the use of internal paternalism in medicine or elsewhere, apart from the more general justification, insofar as it is a justification, that internal paternalism may promote a person's autonomy. For all I have said, the risks associated with allowing internal paternalism in medicine, as well as in other contexts, may outweigh its benefits. Nevertheless, I have argued that it is likely that internal paternalism occurs in certain contexts and that its differentiation from external paternalism is normatively significant. Given the fact that internal paternalism is especially likely to occur in medical practice, physicians should be made aware of its possibility and normative implications, but also of the challenges associated with trying to identify a patient's true or ideal self, as well as of the differences between internal paternalism and forms of paternalism that may rely more heavily on medical knowledge and medical values.

Acknowledgments

I owe my deepest gratitude to Jason Hanna for multiple rounds of invaluable feedback on the paper. I would like to thank audiences at the following venues—and Jay Carlson, Luciana Garbayo, and Gerard Vong in particular—for their feedback: the departments of philosophy at the University of Glasgow and Bar-Ilan University, the 2022 ASBH Philosophy Affinity Group Webinar, the 2021 APA Eastern Division Meeting, and the 13th Annual Rocky Mountain Ethics Congress. I am also grateful to the faculty and graduate students of the Practical Philosophy Workshop at UIUC for excellent comments. Finally, I would like to thank three anonymous referees for the *Canadian Journal of Philosophy*, as well as the handling editor, for their feedback.

References

- Brudney, Daniel. 2009. "Choosing for Another: Beyond Autonomy and Best Interests." *The Hastings Center Report* 39 (2): 31–37.
- Brudney, Daniel, and John Lantos. 2011. "Agency and Authenticity: Which Value Grounds Patient Choice?" *Theoretical Medicine and Bioethics* 32 (4): 217–27.
- Buchanan, Allen. 1978. "Medical Paternalism." *Philosophy and Public Affairs* 7 (4): 370–90.
- Cholbi, Michael. 2017. "Paternalism and our Rational Powers." *Mind* 126 (501): 123–53.
- Conly, Sarah. 2013. *Against Autonomy*. New York: Cambridge University Press.
- Davis, John K. 2004. "Precedent Autonomy and Subsequent Consent." *Ethical Theory and Moral Practice* 7 (3): 267–91.
- Dresser, Rebecca. 1995. "Dworkin on Dementia: Elegant Theory, Questionable Policy." *The Hastings Center Report* 25 (6): 32–38.
- Dworkin, Gerald. 1972. "Paternalism." *The Monist* 56 (1): 64–84.
- Dworkin, Gerald. 1988a. "Autonomy and Informed Consent." In *The Theory and Practice of Autonomy*, 100–20. Cambridge: Cambridge University Press.
- Dworkin, Gerald. 1988b. "Paternalism: Some Second Thoughts." In *The Theory and Practice of Autonomy*, 121–29. Cambridge: Cambridge University Press.
- Dworkin, Gerald. 1988c. "The Nature of Autonomy." In *The Theory and Practice of Autonomy*, 3–20. Cambridge: Cambridge University Press.
- Dworkin, Gerald. 2015. "Paternalism." In *The Cambridge Dictionary of Philosophy*, 3rd ed., edited by Robert Audi and Paul Audi, 762–63. New York: Cambridge University Press.
- Dworkin, Gerald. 2020. "Paternalism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. <https://plato.stanford.edu/entries/paternalism/>.
- Dworkin, Ronald. 1993. *Life's Dominion*. Cambridge, MA: Harvard University Press.

- Enoch, David. 2017. "Hypothetical Consent and the Value(s) of Autonomy." *Ethics* 128 (1): 6–36.
- Feinberg, Joel. 1971. "Legal Paternalism." *Canadian Journal of Philosophy* 1 (1): 105–24.
- Feinberg, Joel. 1986. *Harm to Self*. New York: Oxford University Press.
- Frankfurt, Harry G. 1971. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68 (1): 5–20.
- Groll, Daniel. 2012. "Paternalism, Respect, and the Will." *Ethics* 122 (4): 692–720.
- Groll, Daniel. 2019. "Paternalism, Disagreements, and the Moral Difference." *American Philosophical Quarterly* 56 (1): 57–70.
- Hanna, Jason. 2018. "Hard and Soft Paternalism." In *The Routledge Handbook of the Philosophy of Paternalism*, edited by Kalle Grill and Jason Hanna, 24–34. New York: Routledge.
- Lewis, David. 1989. "Dispositional Theories of Value." *Proceedings of the Aristotelian Society* 63 (suppl. vol.): 113–37.
- MacIntyre, Alasdair. 1981. *After Virtue*. South Bend, IN: University of Notre Dame Press.
- McDowell, John. 1998. "Might There Be External Reasons?" In *Mind, Value and Reality*, 95–111. Cambridge, MA: Harvard University Press.
- Meyers, Diana T. 1989. *Self, Society, and Personal Choice*. New York: Columbia University Press.
- Mill, John Stuart. 1989. "On Liberty." In *On Liberty and Other Writings*, edited by Stefan Collini, 1–15. Cambridge: Cambridge University Press.
- Pope, Thaddeus. 2004. "Counting the Dragon's Teeth and Claws: The Definition of Hard Paternalism." *Georgia State University Law Review* 20 (3): 659–722.
- Quong, Jonathan. 2011. *Liberalism without Perfection*. Oxford: Oxford University Press.

- Savulescu, Julian. 1995. "Rational Non-Interventional Paternalism: Why Doctors Ought to Make Judgments of What Is Best for Their Patients." *Journal of Medical Ethics* 21 (6): 327–31.
- Scoccia, Danny. 2018. "The Concept of Paternalism." In *The Routledge Handbook of the Philosophy of Paternalism*, edited by Kalle Grill and Jason Hanna, 23–11. New York: Routledge.
- Shiffrin, Seana Valentine. 2000. "Paternalism, Unconscionability Doctrine, and Accommodation." *Philosophy & Public Affairs* 29 (3): 205–50.
- Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127 (1): 109–66.
- Street Sharon. 2008. "Constructivism about Reasons." In *Oxford Studies in Metaethics*, vol. 3, edited by Russ Shafer-Landau, 207–45. New York: Oxford University Press.
- Street, Sharon. 2010. "What Is Constructivism in Ethics and Metaethics?" *Philosophy Compass* 5 (5): 363–84.
- Taylor, Charles. 1991. *The Ethics of Authenticity*. Cambridge, MA: Harvard University Press.
- Thaler, Richard H., and Cass R. Sunstein. 2009. *Nudge*. New York: Penguin Books.
- Williams, Bernard. 1981. "Internal and External Reasons." In *Moral Luck: Philosophical Papers 1973–1980*, 101–13. Cambridge: Cambridge University Press.
- Williams, Bernard. 1995. "Internal Reasons and the Obscurity of Blame." In *Making Sense of Humanity and Other Philosophical Papers 1982–1993*, 35–45. Cambridge: Cambridge University Press.
- Williams, Bernard. 2001. "Postscript." In *Varieties of Practical Reasoning*, edited by Elijah Millgram, 91–97. Cambridge, MA: MIT Press.