

CHAPTER 6

Kantian paternalism and suicide intervention

Michael Cholbi

Being theoretically opposed to paternalism is easy. That individuals, groups, and governments in particular, should refrain from interfering with individuals' liberty, except in order to prevent exercises of liberty that harm others or infringe their rights, is an attractive philosophical proposition.

However, anti-paternalism is a bit tougher in practice. For example, many liberal-minded individuals who think of themselves as having anti-paternalist *bona fides* nevertheless appear to endorse various interferences with individual liberty. Mandating the wearing of seatbelts, restricting tobacco advertising, requiring minimal levels of health insurance: All of these might be justified (in part) on the grounds that they result in improvements in individual well-being. Wearing our seatbelts, refraining from smoking, and having minimal access to affordable health care are good for us. But how can support for these policies be consistent with opposition to paternalism?

The answer I favor is that the origins of our anti-paternalistic sentiments are as much Kantian as libertarian. For the Kantian, it is not liberty as such, the mere exercise of one's will, that paternalism threatens. Liberty instead matters because of, and to the extent that, exercises of liberty are exercises of *rational autonomy*. The autonomy in question cannot be reduced either to our ability to act on our choices or to our ability to choose what to do.¹ Human action, on the Kantian picture, should not be seen simply as physiologically instigated bodily movement. We act for reasons, and our capacity to do so is a capacity for rational self-governance or self-direction. For the Kantian, paternalism is ethically suspect not because it constrains our behavior, but because it constrains our actions, and since to act is to act for a reason, paternalism can interfere with our

I gratefully acknowledge the following for their helpful feedback and input on this article: David Adams, Larry Alexander, Richard Arneson, Christian Coons, David Faraci, Kalle Grill, Doug Husak, Kate Manne, Peter Ross, Thomas Schlamme, Danny Scoccia, and Dale Turner.

¹ Husak, "Paternalism and Autonomy," 35–38, notes these other understandings of "autonomy."

rational autonomy. More precisely, paternalism is worrisome because it makes the object of paternalism a “means to ends of the paternalist’s making, rather than being recognized as the source of his or her own ends.”²

But notice that by making liberty the handmaid of rational autonomy, the Kantian opens the door to justifiable paternalism. For interferences with individual liberty do not necessarily violate or undermine rational autonomy. Some interferences with individual liberty can help us achieve our rationally chosen ends instead of thwarting them,³ either by advancing these ends directly or enhancing our capacity to realize these ends. And since we are prone to inattention, poor planning, weakness of will, and the like, we have good reason to assent to interferences with our liberty that preserve or enhance our autonomy or that advance the ends we have rationally chosen. I do not wish to die because I did not wear my seatbelt, or because I became addicted to tobacco, or because I lacked minimal access to health care. Constraining my liberty so as to reduce the likelihood of these outcomes is a welcome sort of paternalism, since it is these very outcomes that I rationally desire for myself. Indeed, a reflexive unwillingness to endorse paternalistic interventions that promote the achievement of my rationally chosen ends raises the question of just how much, *ceteris paribus*, I value the ends that are thereby promoted. If I were to reject such paternalistic interventions, I would in effect be opting to harm myself, not in the sense that I endorse my being injured, but in the sense that I endorse my being worse off with respect to the fulfillment of my rationally chosen ends. I would therefore be instrumentally irrational not to endorse such interventions. For absent such interventions, I may choose courses of action that either make the fulfillment of my ends less likely or that frustrate my pursuit of my chosen ends because the means I have chosen to fulfill those ends are insufficient to realize them. (I may not buy minimal health insurance, say, thereby placing myself at risk of incurring a medical condition I wish to avoid either for its own sake or because it hampers my ability to pursue my other chosen ends).

I make no pretense of suggesting that this position represents Kant’s own position on paternalism. Indeed, I suspect Kant would be ambivalent about it.⁴ Nevertheless, this *Kantian paternalism* (KP)⁵ is an attractive view,

² Kleinig, *Paternalism*, 38.

³ Dworkin, “Paternalism” (1972).

⁴ See for instance Kant’s remarks in his “Theory and Practice” concerning governments that operate on a “principle of benevolence” towards citizens. See Rosen, *Kant’s Theory of Justice*, 15–16, 88–89, 176, for discussion of Kant’s position.

⁵ The Kantian pedigree of this position does not restrict its appeal to those with otherwise non-Kantian sympathies. Kleinig, for example, defends a view of paternalism resting on “autarchy,” or “the developed capacity for rational choice” (*Paternalism*, 21), but appeals more to Mill than to Kant.

to my eyes, not only because of the limited scope of the interferences it permits, but also because it captures the core sentiment behind the anti-paternalistic liberal tradition. We may not interfere with an individual's rationally chosen ends – and in fact, must not pretend to know better which ends individuals ought to pursue – but we may interfere with her irrational pursuit of those ends. KP therefore acknowledges what I take to be the core objection to paternalism: Paternalism, because it involves the “use of coercion to achieve a good which is not recognized as such by those persons for whom the good is intended,”⁶ amounts to substituting others' judgment concerning what a person's good is for that person's own judgment concerning her good,⁷ thereby failing to respect the individual as a locus of rational agency. Paternalism, on this Kantian view, is instead justified when it advances the good of individuals *as they conceive of that good*. So KP is wholly neutral with respect to conceptions of the good: A person whose good is to count blades of grass is no more susceptible to paternalistic interference on this view than a person whose good is to save the rainforests. Thus, by rooting the opposition to paternalism in rational autonomy instead of liberty, KP substantiates the intuition that wrongful paternalism interferes with what we rationally seek, not just with what we do.

This is only a sketch of a view about paternalism. Here I want to extend and then to problematize this sketch. First, I wish to defend KP, not so much by means of direct argument but simply by carefully articulating its commitments and contrasting it (favorably, I hope) with other views about justifiable paternalism. I will argue that KP is well situated to, as Feinberg put it, “reconcile our general repugnance for paternalism” with the “seeming reasonableness” of some paternalistic measures.⁸

It would of course be difficult for me to explore KP's practical implications in any depth here. Indeed, the question of paternalistic *regulations* will turn out to be complex. For example, whether KP would justify legal requirements to wear seatbelts, restrictions on tobacco advertising, or health insurance mandates would depend on contingent facts about the diverse conceptions of the good found within a particular political community, as well as philosophical questions about majoritarian rule, political rights, etc. Hence, I will sidestep these public policy matters to focus on an application of KP within the realm of personal and professional ethics,

⁶ Dworkin, “Paternalism” (1972), 68. Also Scoccia, “Paternalism and Respect for Autonomy,” 320.

⁷ Shiffrin, “Paternalism, Unconscionability Doctrine, and Accommodation,” 220.

⁸ Feinberg, *Harm to Self*, 25.

namely, the conditions under which KP licenses intervention in the suicidal plans of others. Given the momentousness of decisions concerning life and death, intervening in suicidal plans or acts is an area where liberals in particular shall want to tread lightly. In particular, if the arguments I make here are correct, there is a class of suicidal plans or acts, those rooted in the condition I will call *suicidal nihilism*, the interference in which is not permitted by KP, a conclusion that highlights KP's crucial presupposition that agents possess ends to which they are rationally committed. KP could only license interference in nihilistic suicidal behavior by adopting what I call a "pro-vitalist" bias that violates KP's ostensible liberal neutrality.

KANTIAN PATERNALISM DESCRIBED

Here is my official statement of Kantian paternalism:

Interference with an individual's liberty for her own sake is justified absent her actual consent only to the extent that such interference stands a reasonable chance of preventing her from exercising her liberty irrationally in light of the rationally chosen ends that constitute her conception of the good. More specifically, interference with an individual's liberty is permissible only if, by interfering, we stand a reasonable chance of preventing that agent from performing actions she chose due to *distorted reasoning* and which would result in that agent's rationally chosen ends not being as fully realized as they would have been had she so acted.

KP thus understands paternalistic interference not as licensed by others' beliefs that an individual is pursuing ends she ought not to pursue. Rather, KP licenses paternalism in order to make more probable the realization of her ends when that realization has been rendered less probable thanks to generic maladies of human rationality. KP thus does not allow others to paternalistically interfere on the grounds that those others know better what is good for an individual. Rather, KP licenses paternalism because others may know better, in some specific circumstances at least, how that individual's good can be achieved. When justified, paternalism acts to ensure not that individuals make *the* best choice but to assist them in pursuing *their* best choice.

I would first like to note three features of KP before proceeding to articulate the heart of this theory, namely, how certain instances of instrumental irrationality justify paternalistic interference.

First, KP allows for paternalistic interference in order to help agents achieve the rationally chosen ends that constitute their conception of the good. It does not therefore license paternalistic interventions which would

enable me to satisfy desires that, from my own perspective, are relatively trivial. I want salmon salad from my favorite restaurant for lunch, but I won't be able to afford it if I buy my usual morning coffee. KP would not suggest that others who know of my lunch preferences should forcibly prevent me from buying my morning coffee by, say, stealing my wallet and returning it to me just prior to the lunch hour. My desire for salmon salad, unlike my desire to live until retirement age, say, is too peripheral to my conception of the good (i.e., my total plan of life, including my central projects or commitments) to be a desire on behalf of which paternalistic intervention is warranted. KP thus does not provide support for interfering with "trifling" liberties whose exercise is only distantly related to the goods we take to centrally define the shape and orientation of our lives.⁹

Second, KP offers a necessary, rather than a sufficient condition, on paternalistic interference because it recognizes that the moral and non-moral costs of such interference can sometimes outweigh its benefits. KP is not a consequentialist position, but nor is it indifferent to the consequences of paternalistic interference. When others interfere in the service of my rationally chosen ends, such interference places demands on their energy, resources, or wealth that may be too great to justify many interferences. Moreover, KP need not deny the *moral* costs of paternalistic interference. Acting autonomously in the light of one's own rationally chosen ends is itself valuable, even independent of the value of the ends subserved by so acting, so this moral cost must be weighed against the value of the rationally chosen ends the achievement of which a paternalistic intervention might further. Indeed, our own worth or self-respect is itself a good realized by autonomous action, and so in some cases, we may rationally prefer to suffer the diminished realization of our conceptions of the good instead of having our rational autonomy curtailed by paternalistic interference.¹⁰ I make no pretense here of identifying when trade-offs between rational autonomy and the realization of one's good are appropriate. But it does not seem necessarily irrational, for instance, to conclude that even if criminalizing tobacco use would often advance many smokers' conception of their good, the pervasiveness of the paternalistic interferences needed to enforce such criminalization would result in infringements of rational autonomy whose disvalue is greater than the value of the enhanced realization of individuals' conception of the good.

⁹ Kleinig, *Paternalism*, 75; Anderson, "What Is the Point of Equality?," 301–302; and de Marneffe, "Avoiding Paternalism," 68.

¹⁰ Scoccia, "Paternalism and Respect for Autonomy," 323.

KP endorses the thesis that paternalistic intervention is warranted when “we are unable to govern our lives consistent with the goals and values that comprise the ‘self’.”¹¹ However, the self on behalf of whom KP licenses paternalism is the self present *in the circumstances and at the time* at which the paternalistic intervention is to occur. This has two implications. First, the interferences KP licenses are bona fide instances of paternalism, not pseudo-paternalistic “interferences” justified by appeal to consent, hypothetical or actual.¹²

Actual consent, on my view, renders intervention morally permissible because when an agent provides her rational consent, we are thereby given decisive evidence that our intervention with her liberty does not interfere with her rational autonomy. Indeed, we could have no better evidence that it does not so interfere, and in an important sense, the presence of such consent makes it such that our intervention is no interference at all.¹³ Furthermore, what licenses paternalism, according to KP, is not that an agent would, hypothetically or in the future when she is more mature, knowledgeable, rational, etc., consent to such treatment. KP admits that such hypothetical or future-oriented consent can be rational. However, it takes this normative fact as derivative from the more basic normative fact that justifiable paternalism – paternalism to advance a person’s conception of the good – is treatment that the individual herself has reasons to accept. Paternalistic interference therefore does not privilege the desires or reasons of some future and presumably more rational or “authentic” self against the less rational present self.

Likewise, KP does not license paternalism by appeal to a “real” past self whose “true” interests have been temporarily occluded.¹⁴ No doubt individuals sometimes undergo abrupt and radical revisions in their conceptions of the good, revisions of character or judgment which raise questions about whether there exists an identity relation across time between the earlier and later individual. We may reasonably wonder to what extent an addict, a schizophrenic, or someone under the spell of a religious cult really is the same individual as they were prior to these conditions. However, under Kantian paternalism, paternalistic interferences must be justifiable to the actual individuals subjected to them, not to some other counterfactual or idealized counterpart of the individual subjected to the act of paternalism.

¹¹ Edwards, “Beyond Mental Competence,” 274.

¹² Dworkin, “Paternalism” (1972) and Rawls, *A Theory of Justice*, 248, are among the best known defenders of hypothetical consent justifications of paternalism. See also Kleinig, *Paternalism*, 59–67, for discussion of ways that paternalism can be justified by appeal to hypothetical, subsequent, etc., consent.

¹³ Husak, “Paternalism and Autonomy,” 30–31.

¹⁴ Edwards, “Beyond Mental Competence” defends such a justification of paternalism. See Kleinig, *Paternalism*, 58–59, for background.

Granted, cases where an individual's conception of the good diverges wildly from the conception of the good she pursued or espoused in the recent past raise complex *epistemic* questions for KP. Such evaluative interruptions complicate outsiders' efforts to determine the content of an individual's conception of the good, and hence, whether their conduct is instrumentally irrational. But KP denies that paternalism is permissible in order to revive a past self or past conception of the good.

INSTRUMENTAL IRRATIONALITY AND OUR NORMATIVE POWERS

Central to KP is the notion that paternalism is warranted when agents exhibit a certain form of instrumental irrationality. Intuitively, there is a difference between an *agent irrationally* choosing inadequate means to her chosen end and an *irrational agent* choosing inadequate means to her chosen end. Suppose that an agent with a particular end possesses all the relevant information or knowledge needed to determine the best means to her end, and she suffers from no cognitive defect, mental disorder, or infirmity of rationality. In such a case, if the agent in question nevertheless chooses a means that thwarts the achievement of her end, KP does not recommend interfering with her choosing that means. For while she chose irrationally, her deliberation, culminating in her choice of means to her end, was not irrationally shaped. The "inputs" to her deliberation (the relevant information about options, etc.) were sufficient for it to be possible for her to have chosen adequate means to her ends. Furthermore, her mechanisms for forming intentions on the basis of these inputs were intact. We can only attribute her failure of instrumental rationality to her: She is a rational agent who chose irrationally. As I shall put it, she made a *mistake* of practical rationality.

In contrast, an agent whose deliberation is *distorted* either deliberates on the basis of false beliefs, insufficient information, etc. she possesses due to cognitive defects, mental disorder, or infirmity of rationality, or her deliberative and intention-forming capacities are warped or distorted by said conditions. The inadequacy of her chosen means to her ends, i.e., the inadequacy of the "outputs" of her deliberation, is traceable to these defects, disorders, etc. With respect to this particular act of choosing, she was an irrational agent. KP thus disallows interference to prevent mistakes of instrumental rationality, but allows interference to prevent errors of instrumental reasoning due to distortions of rationality.¹⁵

¹⁵ Rawls, *A Theory of Justice*, 249, appears to endorse the general line of reasoning outlined here. Note that the instrumental irrationality that can justify paternalism cannot be equated with a lack of

As an abundance of literature from behavioral economics indicates, the former sort of irrationality (i.e., simply erring in determining how best to pursue our rationally chosen ends) is such a pervasive feature of the human condition that were paternalistic intervention licensed in such cases, it would eviscerate our sphere of autonomous action. Moreover, we have contractualist-type reasons for rejecting paternalistic interferences when we are imperfectly rational even though we have reasons not to reject such paternalistic interferences when our agency is rationally distorted. In caring about our rational autonomy, the object of our concern is a capacity which enables us to choose our ends, and the means to such ends, in accordance with reason. But we do not simply seek to act on the best reasons. We seek to *choose* to act from a recognition of these reasons. Mental disorders, weakness of will, and other infirmities of rationality lead us not to act on what we would take to be the best reasons for action were it not for the influence of these infirmities.

The Kantian paternalist arrives at this position by reflection on our *normative powers*. Human agents are endowed with two such powers. The first, echoing Rawls, is the capacity to fashion a conception of the good, a person's vision (however rough) of her primary aims, convictions, and concerns, the pursuit and achievement of which the person takes to make her life meaningful. The Kantian paternalist readily admits that some conceptions of the good are unreasonable with respect to living on cooperative terms with others or unreasonable *tout court*. Hence, this normative power can be badly exercised. Nevertheless, this normative power so centrally defines our agency that paternalistic interference with it would impugn our rational autonomy. KP thus adopts a stance of neutrality with respect to conceptions of the good and denies that paternalistic actions whose purpose is to impose (or compel the adoption of) a better conception of the good is ever permissible.

The other main normative power is the power of *instrumental rationality*, the power to determine the most effective means to the ends that constitute our conception of the good. By "means" here, I have in mind both causes (in the sense that, for instance, selecting a particular career may be a cause of an effect – income – that is among the elements of one's conception of good) and constituents or specifications (in the sense that selecting a career

deliberation or forethought. Though careful deliberation may tend to produce instrumentally rational choices, it does not follow that hastily made choices are irrational. Also, some agents, having concluded that impulsive choices tend to be more rational for them than those made with great care, adopt impulsivity as a kind of rational strategy. See Scoccia, "Paternalism and Respect for Autonomy," 321.

as a firefighter may instantiate an element of one's conception of the good, such as valuing public service). This normative power plays a subsidiary role vis-à-vis the former. Whether the power of instrumental rationality is successfully exercised depends on an agent having a specific conception of the good on the behalf of which this power is exercised. Here too, this normative power can be exercised well or badly. Indeed, this must be the case if instrumental rationality is a normative power. Normativity is built into the exercise of reason, so it must be possible for us to succeed or to err in its exercise. But even when exercised badly, this power is still exercised and is worthy of others' respect. KP thus denies that paternalism is warranted to correct for actual or anticipated irrationality when we make a mistake of rationality. To permit paternalistic interference in those circumstances would show little respect for instrumental rationality *qua* normative power. But in circumstances when we may irrationally pursue our conceptions of the good due to distortions of irrationality – when, in some crucial sense, we cannot exercise this power successfully – KP understands paternalistic interference not as an interference with an exercise of a normative power. Paternalistic interference instead protects agents from harming themselves by acts resulting from causes that circumvent their normative power of instrumental irrationality.

That instrumental irrationality can be the product of mistake or of distortion is, I hope, a clear enough notion. Still, I anticipate that in some instances, it will be controversial which of these are the source of instrumental irrationality. Take weakness of will. Shall we say that weakness of will constitutes an infirmity of rationality, so that if the weak-willed person is instrumentally irrational in the pursuit of her conception of the good, her irrationality stems from a distortion? Or is weakness of will itself a kind of rational mistake, say, a failure to keep in view the relevant evidence concerning how an act serves one's good? And what if weakness of will is habitual, so that a person could have prevented being weak-willed if she had acted differently in the past? These are hard questions, but the haziness of the boundary between mistakes and distortions should not lead us to believe there is no boundary with clear cases on either side. In some cases of instrumental irrationality, only intricate philosophical argumentation, along with an abundance of empirical evidence, can determine if it results from distortion or mistake.

In other instances, whether instrumental irrationality results from mistake or distortion may be epistemically opaque to outsiders. Consider Mill's famous example of the heedless bridge crosser:

If either a public officer or any one else 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.¹⁶

KP concurs, in general outline, with Mill's conclusions regarding this kind of case. The central issue is whether the heedless bridge crosser, whose crossing of the rickety bridge would be irrational, crosses the bridge due to a mistake of rationality or a distortion of her rationality. Here bystanders (and policy-makers) are presented with an epistemic challenge. For as Mill says, no one but the bridge crosser is in a position to judge "the sufficiency of the motive" (i.e., the prudential rationality) of crossing the bridge. We may have strong evidence that the decision stems from distortion when the bridge crosser is a child or delirious, etc. But in many cases, we will lack the evidence necessary to know whether paternalism is warranted. A compromise thus seems warranted: We presume that paternalistic intervention is reasonable but should be modest – a warning, as Mill recommends, rather than a forcible attempt to prevent the bridge crosser from crossing over. Neither those bridge crossers who irrationally cross the bridge due to mistake nor those who irrationally cross the bridge due to distortion have a reasonable basis for complaining that their normative powers, or their rational autonomy, were unduly interfered with under this policy.

Hence, according to KP, what triggers justifiable paternalism is not the nature of the good a person pursues but the exogenous imprudence with which she pursues it. As Kant remarked in his elaboration of the hypothetical imperative, to will an end is to will the means to that end. Prudential rationality, then, is not a substantive value or good, but a formal demand of practical rationality, a rationality whose functioning can, thanks to various all-too-human infirmities, become distorted so that our choices and actions do not succeed in realizing our conception of the good.¹⁷ KP thus rejects Richard Arneson's contention that instrumental rationality is "a value which we have no more reason to impose on an adult against his

¹⁶ Mill, *On Liberty*, chap. 5, para. 5. ¹⁷ Scoccia, "Paternalism and Respect for Autonomy," 320.

will for his own good than we have reason to impose any other value on paternalistic grounds.”¹⁸ If I am correct, then we have such reasons to “impose” instrumental rationality on agents when they are instrumentally irrational due to distorted rationality, and agents themselves have reasons to accept this imposition, reasons stemming from their commitment to the realization of their conceptions of the good. But KP remains steadfast in its neutrality amongst conceptions of the good.

To summarize so far: KP provides a credible basis for opposing paternalism as wrongful interference with our rational autonomy, while also allowing for a modest spectrum of paternalistic interventions that respect our rational autonomy by advancing our rationally chosen ends or conceptions of the good. More exactly, it permits such interventions when we are instrumentally irrational due to distortions of rationality, but not when such irrationality rests on rational error. Such a position captures the central intuition behind opposition to paternalism, namely, that it supplants an agent’s view of her own good with that of another, and in so doing, “shows significant disrespect for those core capacities or powers of the agent that underwrite and characterize his autonomous agency.”¹⁹

THE IRRATIONALITY OF SUICIDE

In my estimation, KP offers an attractive overall stance on paternalism, a version of soft paternalism rooted in the value of rational autonomy. Nevertheless, its limitations become apparent when we consider how it might justify paternalistic interventions to prevent suicide. If I am correct, then in those cases in which we most desire to intervene and find suicide most tragic, KP bars us from paternalistic interferences.

The chief difference between how a libertarian anti-paternalist and a Kantian paternalist approach the ethics of suicide intervention is that, for the latter, questions of the rationality of suicidal conduct will loom large. For KP, the central question to be asked about suicide is whether the suicidal agent has rationally determined that suicide fulfills her chosen ends, whatever those ends happen to be, and if not, what the source of her irrationality is.

Some philosophers offer *a priori* grounds for skepticism about rational suicide. Philip Devine, for instance, argues that in order for suicide to be rational, we must know what death is like, but since that is presumably

¹⁸ Arneson, “Mill Versus Paternalism,” 474.

¹⁹ Shiffrin, “Paternalism, Unconscionability Doctrine, and Accommodation,” 220.

unknowable at the time when suicide is chosen, the conditions for suicide being rational cannot be met.²⁰ In a similar vein, Christopher Cowley proposes that rationality is essentially future-oriented, but since the act of suicide is the decision not to have a future, it cannot be rational.²¹

If sound, these arguments would show that the notion of a “rational suicide” is a conceptual *non sequitur*. But they are clearly not sound. Such arguments presuppose the Epicurean notion that the rationality of ending one’s life turns on comparing the life one would have by continuing to live with the state of being dead. On the assumption that being dead is not a state of a person – or at least not an *experiential* state of a person – the rationality of suicide becomes unintelligible. A state cannot be compared with a non-state. But as Derek Parfit and others have convincingly argued, this “two-state requirement” should be rejected. Instead, the comparison relevant to considering whether suicide is rational is between two lives or two life-courses, one in which a person’s life comes to a premature end, say, at time t , with the life she would have (or probably would have had) had she lived until a later time $t+$. In some cases at least, a person is clearly better off if her life ends at t rather at $t+$. To appreciate this, observe that a person whose life is saved in perilous circumstances can clearly benefit from her life being extended, but this benefit is not cogently explained by saying that she is better off being alive than being dead. Rather, her longer life was a better life than the shorter life she would have ended up with had she not been saved from peril. But if a person can be benefitted by continuing to live, then presumably a person could in principle be benefitted by shortening her life, so long as the shortened life is better for her than the longer life. On this *comparativist* account, suicide can benefit a person, and hence be rational, if (on balance) it prevents her from suffering various harms that she would otherwise have endured.²²

Hence, we cannot reject the possibility of rational suicide on *a priori* grounds alone. Nevertheless, that suicide is not in principle irrational does not tell us what conditions a suicide must meet to be irrational, and more specifically, irrational in a way that warrants paternalistic intervention. Philosophers, psychologists, and others have offered a number of characterizations of rational suicide.²³ Wading into the many intricate issues raised by this literature is orthogonal to my purposes here. For the most part, these

²⁰ Devine, “On Choosing Death.” ²¹ Cowley, “Suicide Is Neither Rational nor Irrational.”

²² See Parfit, *Reasons and Persons*, 487–489; Luper, *The Philosophy of Death*, 82–88; and Bradley, *Well-Being and Death*, 47–52, for defenses of this comparativist thesis.

²³ See, among many, Motto, “The Right to Suicide”; Brandt, “The Morality and Rationality of Suicide”; Graber, “The Rationality of Suicide”; Werth and Cobia, “Empirically Based Criteria for

characterizations agree that suicide is rational if and only if it meets two broad sets of criteria. The first set of criteria is *cognitive*, relating to the process and context by which the attitudes of the suicidal individual that inform her decision to end her life are shaped – that her decision not rest on disorder-induced delusions, that she possess minimal competence at causal and inferential reasoning, etc. The second set of criteria are *interest-based*, relating to how well dying would serve an individual's considered interests, etc.

Here my concern is not with the rationality of suicide as such, but with the conditions of rationality suicide must meet in order to preclude the sort of paternalistic intervention that KP allows. As I have already suggested, such intervention is not warranted unless a suicidal agent opts to die when her so opting is instrumentally irrational due to distortions of rationality rather than to mistakes of rationality. In this respect, KP concurs that a rational suicide must meet both cognitive and interest-based conditions. However, this concurrence comes with crucial caveats. Some characterizations of rational suicide treat an individual's interests in a realist manner, as if what constitutes an individual's interests is wholly independent of her actual attitudes. Take, for instance, the oft-cited condition that a suicide is rational only if the person suffers from a condition that is terminal and/or "unremittingly hopeless."²⁴ If "hopeless" here refers not to whether an individual finds her life hopeless but to some objective fact about the hopelessness of her situation, then this requirement encodes a substantive conception of what makes a life worth living. Irrespective of whether this value judgment is *true*, or whether a suicide undertaken in the absence of a terminal or hopeless condition is in some objective sense rational or "reasonable," KP maintains that this requirement not being satisfied is not a mark of an irrational suicide for which outside intervention is morally permitted. KP allows, for instance, that an individual with a non-terminal or treatable condition may nevertheless prefer suicide to being treated, and it would be wrong for us to intervene in her suicidal acts if she meets the other cognitive conditions for rationality. From the standpoint of intervention, what she ought to want or what it is rational to want is not relevant under KP. For KP, intervention hinges on the rationality of the suicidal *person*, not the rationality of the suicidal *act*.

KP is similarly skeptical about a requirement that the "motivational basis" of a suicidal person's decision to die be "understandable to the

Rational Suicide"; Battin, "Can Suicide Be Rational?"; Prado, *Choosing to Die*, 64; Pilpel and Amsel, "What Is Wrong with Rational Suicide"; and Cholbi, *Suicide*, 88–96.

²⁴ Werth and Cobia, "Empirically Based Criteria for Rational Suicide," 233.

majority of uninvolved observers from his (or her) community or social group”²⁵ or that “choosing to die must be prompted by motivation that cultural peers and members of other cultures judge does not unduly override the interest in survival.”²⁶ KP’s neutrality concerning conceptions of the good, and its consequent neutrality about the reasonableness of the ends suicide may serve, entails that societal consensus or approval is irrelevant to whether or not intervention to prevent suicide is morally permissible. As Jerome Motto writes, “What may be an inconvenience, a source of discomfort, or an embarrassment to one person represents unbearable agony, excruciating pain, or intolerable humiliation to another.”²⁷ KP honors the diversity of conceptions of the good that individuals endorse, and hence rejects as a condition of rational suicide that the suicidal act be motivationally intelligible to others.

That being said, KP provides reasonably broad scope for suicide intervention. The central issues for KP are (a) whether an act of suicide is instrumentally irrational given the agent’s conception of the good, and (b) if so, whether this irrationality is a product of rational distortion as opposed to rational error. In other words, the Kantian paternalist must ask: Has an agent reached her decision to end her life via an undistorted rational engagement with the question posed earlier, namely, is the life course in which she ends her life via suicide at time t better by her lights than her continuing to live to times t_1 , t_2 , etc.?

There is sufficient evidence about the psychology of suicidal persons to suggest that the typical answer is “no.” The difficulties of postmortem psychiatric inquiry make precise conclusions elusive, but by most accounts, 70 to 95 percent of suicidal persons suffer from mental illness, most commonly affective disorders such as major depression, bipolar disorder, borderline personality disorder, and panic disorder.²⁸ Unfortunately, these statistical findings are not as clear cut as we might like. For in some cases, suicidal thoughts or behavior are both partially explained by the presence of such illnesses while also appearing in the diagnostic criteria for these illnesses, an instance of worrisome explanatory circularity. Furthermore, we cannot straightforwardly infer from the presence of such disorders that their sufferers who seek to die are instrumentally irrational in a way that, according to KP, triggers justifiable suicide intervention.

²⁵ Siegel, “Psychosocial Aspects of Rational Suicide,” 407.

²⁶ Prado, *Choosing to Die*, 46.

²⁷ Werth, *Rational Suicide?*, xi.

²⁸ See Jamison, *Night Falls Fast*, 100–112; Cavanagh et al., “Psychological Autopsy Studies of Suicide”; Joiner, *Why People Die by Suicide*, 192–202.

Mental illness varies in its severity, and few of those with affective disorders are criminally insane. Yet these illnesses are likely to introduce various distortions into the practical reasoning of the suicidal. For although these illnesses do not produce delusions, they tend to result in irrational patterns of thinking and reasoning, patterns that lead agents to wrongly believe that suicide in fact serves their conceptions of the good.

Depression, for example, tends to introduce the kind of instrumental irrationality that would trigger intervention under KP. Richard Brandt has written:

The person who is contemplating suicide is obviously making a choice between future world-courses: the world-course that includes his demise, say, an hour from now, and several possible ones that contain his demise at a later point . . . The basic question a person must answer in order to determine which world-course is best or rational for him to choose, is which he *would* choose under conditions of optimal use of information, when *all* of his desires are taken into account. It is not just a question of what we prefer *now*, with some clarification of all the possibilities being considered. Our preferences change, and the preferences of tomorrow are just as legitimately taken into account in deciding what to do now as the preferences of today.²⁹

But depression, Brandt observes, “primitivizes” our deliberation so that one’s future preferences and values become discounted even more steeply than normal. To end one’s life rationally, one must not only think about one’s present desires but about the desires one would have if one continued to live. And this projection of one’s future self is something that depressed individuals, with their fixation on current suffering, are often unable to do. Thus, depressed individuals are likely to engage the central questions surrounding the rationality of their own suicide in distorted, and hence instrumentally irrational, ways.

Given these considerations, it is therefore probable that any given suicidal act is rationally distorted, so Kantian paternalists should adopt suicide intervention as a defeasible or *prima facie* principle. The case for such paternalistic intervention is strengthened by the fact that if suicide is instrumentally irrational and would not advance the individual’s rationally chosen ends, then suicide would (obviously!) be profoundly destructive to their conceptions of the good, regardless of what those conceptions happen to be. In other words, the costs of failing to engage in justified paternalistic intervention are much greater in the case of suicide than with respect to other irrational or self-destructive behaviors. Compare instrumentally

²⁹ Brandt, “The Morality and Rationality of Suicide,” 66.

irrational suicide to another common instance of human instrumental irrationality: overeating. Gluttony is a common human vice, and a case could be made that at least sometimes, overeating is the product of all too common distortions of practical reasoning (situational rationalization, the use of food as self-medication, etc.). Moreover, overeating is often irrational in light of our conceptions of the good, since overheating can lead to long-term harms such as premature death or immobility, outcomes that are at odds with conceptions of the good many of us share. But there is a weaker case (on paternalistic grounds) to interfere in others' dietary habits than there is to interfere in others' suicidal acts, even though, according to KP, such interference could be justified in either instance. This is because the harms of overeating are importantly different from the harms of suicide. In comparison with obesity, for instance, the harms are severe, irreversible, and imminent.

In concert, these two considerations – that many suicidal persons choose to die on the basis of distorted instrumental irrationality and that such choices are prudentially more costly than other instrumentally irrational choices – provide the basis for a defeasible permission to paternalistically interfere in the suicidal plans or acts of others.

KANTIAN PATERNALISM AT ITS LIMITS: SUICIDAL NIHILISM

Yet KP rests on an important psychological assumption, namely, that individuals have conceptions of the good to which they are committed and which motivate their actions. When morally justified, paternalistic interference alters the individual's situation in the expectation that such alteration will result in a better future when seen from her present standpoint on her conception of the good.

But there is a class of suicides for whom these background assumptions are false. Many suicidal individuals, especially the depressed, are *pessimistic*. Their despair stems from the belief, perhaps false, that their conception of the good stands a very low probability of being realized. Either the world is inherently hostile to their ends, they feel constitutionally incapable of realizing these ends, or both. For the pessimist, her situation is hopeless, but only because she perceives that her pursuit of the ends to which she is committed is futile.

But other suicidal individuals are hopeless not because they believe the pursuit of happiness is futile, but because there no longer exists a state they recognize as "their happiness." In 1793, Maria von Herbert, in one of the most audacious philosophical performances ever, wrote to Kant describing the state of mind that had led her to contemplate suicide:

My vision is clear now. I feel that a vast emptiness extends inside me, and all around me – so that I almost find myself to be superfluous, unnecessary. Nothing attracts me. I'm tormented by a boredom that makes life intolerable. Don't think me arrogant, but the demands of morality are too easy for me . . . You can see, perhaps, why I only want one thing, namely to shorten this pointless life, a life which I am convinced will get neither better nor worse.³⁰

At this point, von Herbert is beyond pessimism. She is instead “tormented” by boredom, living a life she finds “superfluous” with nothing to attract her. Her desire for suicide is not motivated by a frustration with realizing her conception of the good. It is instead motivated by the realization that she no longer has a conception of the good. Von Herbert exhibits suicidal *nihilism*, not simply pessimism.³¹

We cannot say precisely how many suicidal individuals are nihilists about the good instead of merely pessimists about the good. But nihilistic sentiments, similar to von Herbert's, are commonly expressed in suicide notes. The journalist Hunter S. Thompson and the Academy Award-winning actor George Sanders both left notes citing boredom as their motivation. Expressions of fatigue, ennui, and inevitability are also common.³² Furthermore, suicidal individuals do not react to the world as if they are committed to the achievements of their personal ends. As suicidal ideation intensifies, concern for one's future, along with the reactive attitudes of frustration, pride, and regret, diminish.³³ Suicide, especially when influenced by depression or other affective disorders, often reflects a disenchantment with the world, a sense of a world drained of value. As Barbara Herman has written, our concept of ourselves is intimately bound up with our sense of being able to shape our future experiences to our desires: “[W]e not only can envisage what it would be like to do one thing *or* another, we recognize that we have a future whose shape we can affect. We come to have a concept of a life – something that will connect our experiences – and about which we feel satisfaction, or not.³⁴ This concept of a life about which we can feel satisfaction or not is precisely what nihilistically suicidal individuals lack.

On its face, nihilistic suicides look like those KP would allow us to intervene in for paternalistic reasons. For the conditions associated with

³⁰ From von Herbert to Kant, January 1793, in Zweig, ed., *Kant: Philosophical Correspondence, 1759–1799*.

³¹ I defend similar claims about suicidal nihilism in my “Suicide Intervention and Non-Ideal Kantian Theory”; “A Kantian Defense of Prudential Suicide”; and “Depression, Listlessness, and Moral Motivation.”

³² Leenaars, *Suicide Notes*.

³³ See the summary of suicidal psychology in Baron, *Essentials of Psychology*, 471–472.

³⁴ Herman, “Transforming Incentives,” sec. 3.

such suicides are among those likely to produce distortions in instrumental rationality. But the question of the rationality of nihilistic suicide is beside the point here. For paternalistic interference requires an assumption that nihilistic suicides do not satisfy. KP licenses paternalism in order to enable individuals to act more prudently in the pursuit of their conceptions of the good. But nihilistic suicide occurs when agents see their lives as superfluous precisely because they no longer have a conception of the good. In cases of suicidal nihilism, the capacity to effectively pursue one's conception of the good is not compromised. The very capacity for a conception of the good is compromised. The nihilistic endorse no comprehensive rational plan of life. Hence, we cannot paternalistically interfere on behalf of a nihilistic person, for there is nothing on behalf of which to interfere. The nihilistic individual is not rationally choosing suicide in the relevant sense, but nor is she irrationally choosing either. A background assumption necessary to render the attribution (and the non-attribution) of rationality to her choice is not met. Having divested themselves of the value of their future, nihilistic individuals are no longer individuals for whom acting prudently, or imprudently, is intelligible.³⁵

The only strategy by which to show that paternalism would be justified in cases of nihilistic suicide would be to appeal to the irrationality of nihilism itself. Life is precious, we often believe, a gift not to be squandered, a journey whose ups should be savored and whose downs we should make every effort to surmount. With this sentiment in mind, we might then imagine ourselves as proxies for suicidal individuals, determining that the inherent value of life warrants our stopping nihilistic suicides for the sake of those who would otherwise end their lives. But to take such a stance violates the neutrality concerning conceptions of the good that lies at the heart of KP. Nihilistic suicide is the product of not having a recognizable conception of the good. However, to rest paternalistic interference on the notion that life is inherently valuable requires attributing a specific, and quite controversial, *pro-vitalist* conception of the good to those who do not share it. We would not be substituting our judgment concerning the meaning or significance of life for the suicidal individual's judgment, since nihilism just is the condition of spurning such judgments. However, we would be subjecting her to an interference on the basis of a value she does not endorse.

It is only by relinquishing KP's neutrality about the value of a person's ends that we could justify paternalistic interference in nihilistic suicides,

³⁵ Scoccia, "Paternalism and Respect for Autonomy," 325.

since the requirement (implicit in KP) that paternalistic interference is justified only in order to advance agents' own good as they understand it is not satisfied in these cases. The only defensible bases for keeping the suicidal nihilistic alive must be non-paternalistic, for instance, that their continuing to live is required by some moral duty or that we prefer they continue to live.

CONCLUSION

I offer these points concerning suicide intervention not to suggest that KP should be rejected. As suggested above, I believe KP is as good as any extant account of paternalism in threading the philosophical needle of explaining what is objectionable about paternalism while still making sense of when paternalism might be permissible. Yet it is puzzling to think that those who may nevertheless benefit most from suicide intervention are among those whose lives we cannot interfere with in order to realize their happiness. In the end, the problems raised by nihilistic suicide are likely to confound not only KP, but any attempt to justify paternalism.