Chapter 9

A *pro tanto* moral case for assisted death

Isra Black

Introduction

This chapter makes a *pro tanto* case for the moral permissibility of assisted death. It is *pro tanto* in the sense that my principal interest is the *all else being equal* moral evaluation of the dimension of assisted death that comprises one individual, who wishes to die with assistance, and another, who either grants or refuses to grant the former's wish and who provides assistance or does not provide assistance respectively. Put another way, this is not a case for assisted death *all things considered*. Specifically, I shall not consider the detail of other-regarding matters relevant to the all things considered moral permissibility of assisted death. In making a *pro tanto* moral case for assisted death only, I have a substantive aim: to set out a universal foundation for the moral permissibility of assisted death, which in turn contributes to discussion of the *all things considered* moral permissibility of assisted death, and which structures (in part) the institutionalisation of assisted death in law or policy.

I argue that, all else being equal, it is morally permissible for an agent to assist an individual to die just when their wish to die is autonomous and assisting their death is what an agent has most reason to do, where the latter is determined by the intrinsic values of the courses of action available to the agent in the individual's regard.[1] More formally, it is, all else being equal, morally permissible for Q to assist P to die if P's decision to die is autonomous and Q has most reason to assist P's death.

---

[1] I am using they/them/their etc in a singular, gender neutral sense.

I reach my conclusion that assisted death is *pro tanto* morally permissible through discussion of Ralph Wedgwood's (2009) account of how intrinsic values—specifically those instantiated by states of affairs—generate reasons for action, that is, how such values explain what an agent *ought* to do (for a state of affairs, at a time). This metaethical view offers a promising starting point for discussion of the morality of assisted death. It is compatible with the best explanation of the value of death for an individual who dies: the deprivation account (Feldman 1992, 1991). Wedgwood's theory attends to non-consequentialist concerns going to the—causal or intentional—role an agent plays in bringing about a state of affairs. Specifically, it is sensitive to the different ways in which an agent may causally contribute to another's death and the intention–foresight distinction, both of which are often thought relevant to the moral permissibility of assisted death. Finally, on a methodological note, I suggest that it is advantageous to foreground discussion of the nature and source of the moral reasons relevant to the permissibility of assisted death.

By way of plan, I shall first treat some preliminary matters. I then outline Wedgwood's view of intrinsic values and reasons for action. I then make the case for the *pro tanto* morality of assisted death. In the conclusion, I set out in brief some factors relevant to the *all things considered* moral permissibility of assisted death.

Preliminaries

It is important to settle some preliminary matters, which should help to clarify the target of my argument and what I hope to achieve.

For brevity, I shall refer to an individual who dies as *P*, and an individual (or individuals) who is (or are) causally involved in *P*'s death as *Q*(s). For simplicity, I shall assume only one *Q* in what follows. I shall often refer to courses of action as *A* and consequences of or the state(s) of affairs resulting from *A* as *S*. *S* includes all relevant facts about the world for time *t*.

By assisted death, I mean voluntary euthanasia or assisted suicide (Black 2020, forthcoming).[2] The concepts of suicide and of assistance are respectively complex, perhaps surprisingly so. I shall have to rely on an intuitive grasp of what it means for *Q* to assist *P*'s suicide. To amount to euthanasia, *Q* must deliberately *cause P*'s death and death must be *good for* P. If we wish to exclude relevant instances of assisted suicide from the concept of euthanasia, some further specification as to causation is necessary. Something along the lines that *Q* must be the *most proximate* cause of *P*'s death should suffice. All else being equal, euthanasia is voluntary when *P* makes an autonomous positive decision—that is, consent or request—in respect of *Q*'s causing *P*'s death; it is involuntary when *P* makes an autonomous negative decision—refusal—in respect of *Q*'s causing *P*'s death. I take it that an individual's decision is autonomous just when they have adequate information in respect of *S*, have capacity to decide what to do in *S*, and are not subject to autonomy-undermining third party influence. All else being equal, euthanasia is non-voluntary when *P* is presently unable to take an autonomous decision in respect of *Q*'s causing *P*'s death and has no prior autonomous decision that would make euthanasia voluntary or involuntary. I shall argue that involuntary euthanasia may be *pro tanto* morally permissible, but that it is usually *pro tanto* morally impermissible to pursue a course of action that involves involuntary killing. My discussion leaves open the morality of non-voluntary euthanasia.

I shall assume a non-absolutist position in respect of the morality of conduct that intentionally causes or contributes to death. There are many defences of the view that assisted death is absolutely wrong (Finnis 1995; Keown 2002), specific replies (Harris 1995), and a literature that engages critically with the underlying normative machinery of absolutist views

---

[2] Many authors, eg (Lewis 2007a; Coggon 2010), use 'assisted dying' to refer to voluntary euthanasia and assisted suicide. I note elsewhere that 'within the public debate on the legalisation of end-of-life decision-making and assistance [in England and Wales], certain organisations and politicians have attempted to redefine assisted dying as (physician) assisted suicide only. Whatever its political merits, this move has not served the ends of conceptual clarity' (Black forthcoming). As such, I prefer the term 'assisted death'.

(McIntyre 2001; Scanlon 2008). There is little to gain from rehearsing these arguments. In my view, non-absolutism is more interesting—it seeks to map the contours of assisted death, some instances of which will be morally permissible and others morally impermissible. Non-absolutism is also more appropriate in the context of a volume that is largely friendly to the moral, legal, and social permissibility of assisted death.

Finally, I shall not directly defend Wedgwood against other accounts of reason for action. I shall attempt to set out Wedgwood's view in sufficient detail so that readers may gauge its plausibility. I also aim to provide indirect support for the argument that intrinsic values generate reasons for action, insofar as I aim to show that it fares well in reflective equilibrium (Daniels 1979). Wedgwood's theory is compatible with the deprivation account of the value of death; it is sensitive to non-consequentialist intuitions; and, importantly for our purposes, it yields a plausible account of the morality of assisted death.

Wedgwood's intrinsic values and reasons for action

What is a reason for action? As Wedgwood explains, '[a] "reason for agent *x* to do act *A* at time *t*" is some fact about *A* (in relation to *x*'s situation at *t*) that plays a certain sort of role in *explaining* what *x ought* to do at *t*' (2009, 321).[3] As the 'ought' signifies, reasons for action, at least in the sense relevant here, are normative: they explain—or justify—what counts in favour of or against a course of action over any other (Raz 1999).[4]

According to Wedgwood, 'all reasons for action are grounded in facts about how courses of action available to the relevant agent at the relevant time are related to *intrinsic*

---

[3] All emphasis in quotations is original, unless stated otherwise.
[4] There is a specific sense 'ought' takes within Wedgwood's theory of reasons for action (2009, 322)—what ought is about and what ought comprises. First, ought is *practical*, in that it relates, in a general all things considered way, to the end or starting points of deliberation about what to do (322). Second, ought is *objective* in that 'what an agent… ought to do is determined by *all* practically significant facts about the agent's situation—regardless of whether the agent in question actually knows those facts, and even of whether the agent is in a position to know these facts' (322).

values' (2009, 323). Of course, more needs to be said to explain this theory. In this section, I attempt to provide an outline of its moving parts in three stages. I start by detailing what Wedgwood means by intrinsic value and the states of affairs that bear intrinsic value. I then explain how Wedgwood establishes the intrinsic values of courses of action using examples relevant to assisted death. Finally, I explain how the intrinsic value of courses of action generate reasons for action.

Intrinsic value and states of affairs

On Wedgwood's view, we can express intrinsic value in terms of *absolute* goodness or value (324). One way to explain that some thing has absolute value is by appeal to its being the fitting object of an *agent neutral* pro-attitude:

> an item *x* has [absolute value] if and only if it is appropriate for *anyone* who adequately considers *x* to have the corresponding… pro-attitude of some fairly straightforward kind [eg *admiration*]. If it is really is appropriate for absolutely *anyone* who adequately considers *x* to have this sort of pro-attitude towards *x*, this pro-attitude must be an essentially *disinterested* pro-attitude, that is, a pro-attitude that does not depend for its appropriateness on the particular relation that the thinker has towards *x* (325).

The bearers of intrinsic values within this theory are *states of affairs* (327), that is, the sets of circumstances that serve as a description of some thing. For example, we might describe the state of affairs that concerned Tony Nicklinson—who challenged the legal prohibitions on assisted death in England and Wales—immediately prior to his death as: aged 51 years, suffering from 'locked-in' syndrome (*R (oao Nicklinson v Ministry of Justice* [2014] UKSC 38). Three important clarifications are necessary.

First, the description of a state of affairs must be sufficiently detailed so as to permit a determination of its intrinsic value:

> the state of affairs of *x*'s having *P* [where *P* is the property that makes it the case that *x* has value *V*] must… be sufficient for that state of affairs having the corresponding value *V*: that is, it is impossible for this state of affairs to exist without having that value *V*. In that sense, this value *V* is an intrinsic feature of this state of affairs (Wedgwood 2009, 328).

The above description of Tony Nicklinson's circumstances seems insufficient to determine its intrinsic value. An evaluatively adequate description might be: man, aged 51 years, paralysed save for the ability to move his head and his eyes, able to communicate via eye blink computer, regarding his life as 'dull, miserable, demeaning, undignified and intolerable', wishing to die (*Nicklinson* [3] (Lord Neuberger)).

Second, the intrinsic values instantiated by states of affairs come in degrees, that is one state of affairs may instantiate intrinsic values to a greater or lesser extent than another (Wedgwood 2009, 329). The upshot of this idea is that 'each of these values generates a *ranking* of states of affairs' (329), that may permit their comparison.

Third, on Wedgwood's theory the states of affairs that bear intrinsic value are *abstract* entities (327, 329). That is, a state of affairs may '*exist* even if it does not *obtain*' (327). This is important for our purposes, since when considering the moral permissibility of assisted death, it is necessary to rank the degrees of intrinsic value of mutually exclusive states of affairs—*P* receives assistance to die or receives no such assistance respectively. For example, had Tony Nicklinson succeeded at first instance he might have received assistance to die; it transpired that he died of pneumonia having refused 'all nutrition, fluids, and medical treatment' (*Nicklinson* [6] (Lord Neuberger)). More generally, as Wedgwood observes, it seems 'of great importance to ethical theory to be able to *compare* states of affairs that do not actually obtain' (2009, 329).

The intrinsic value of courses of action involving assisted death

The states of affairs relevant to reasons for action (and our discussion of the moral permissibility of assisted death) are *courses of action*, broadly construed to include omissions as well as acts, and sufficiently detailed such that the relevant values are an intrinsic feature of the agent's conduct. A general way of describing such a detailed course of action is: 'acting in

such a way that a state of affairs that instantiates value V to degree d_1 results' (333). Because courses of action are complex states of affairs, that is, made up of other states of affairs, the degree of intrinsic value instantiated in a course of action is complex—is a function of the intrinsic value of the other intrinsically valuable states of affairs the course of action comprises (332).

Two elements determine the intrinsic values of courses of action on Wedgwood's theory. First, there is the intrinsic value of the individuated states of affairs that count as a consequence of a course of action (333-334). Second, there is the degree of agential involvement in bringing about each of the consequences of a course of action (334-336).

The intrinsic value of the each of the consequences of a course of action is determined *comparatively* (333). The explanation of how such a comparison might take place is somewhat complex, but the idea is intuitive. For the set of alternative courses of action available to an agent at a time, there exists a *benchmark of comparison* for each of the consequences that exemplifies an intrinsic value:

> For every agent and every time when the agent is capable of acting, and for every intrinsic value *V* that is exemplified by some of the consequences of courses of action that are available to that agent at that time, there is a relevant "benchmark of comparison" for *V* with respect to the situation of the agent at that time (333).

One way to determine this benchmark of comparison, for example, is to average the degree to which each of the consequences of the alternative courses of action exemplifies an intrinsic value (333). I shall proceed on this basis. Once we have established the benchmark of comparison, the positive or negative degree of intrinsic value of a consequence of a course of action is a function of its situation above or below the benchmark respectively:

> When a consequence of a course of action is *superior* to the benchmark (with respect to the relevant value), the consequence exemplifies this value to a *positive* degree: in other words, having this consequence counts as a *good* feature of the course of action. When a consequence of a course of action is *inferior* to the benchmark (with respect to that value), the consequence

exemplifies this value to a *negative* degree; and that counts as a *bad* feature of the course of action (333).

The upshot of measuring the intrinsic value of the consequences of courses of action comparatively is that 'there is a *separate* benchmark for every practical situation of an agent at a time [and] no unique zero-point that divides *all* states of affairs that instantiate the relevant value' (341).

To make the discussion a little more concrete, imagine that there are two courses of action available to Q: A_1 and A_2. A_1 involves Q assisting P to die, with the consequences that $S_{1.1}$: P's wish to die is fulfilled; and $S_{1.2}$: P suffers for t_1 but avoids suffering for t_2.[5] A_2 involves Q not assisting P to die, with the consequences that $S_{2.1}$: P's wish to die is fulfilled by refusing life-prolonging treatment; and $S_{2.2}$: P suffers for t_1 and t_2. I stipulate that the exercise of autonomy instantiates the intrinsic value of well-being to some degree—I shall discuss the intrinsic values relevant to assisted death shortly. So $S_{1.1}$ and $S_{2.1}$ instantiate the value of well-being to some degree. In both states of affairs P's wish to die is fulfilled, but in $S_{1.1}$, P's exercise of autonomy extends, through the receipt of assisted death, to deciding how and when to die, whereas in $S_{1.2}$, P's exercise of autonomous refusal of life-prolonging treatment cannot control the manner of their death or its exact timing. If the benchmark of comparison is the average of the degree to which $S_{1.1}$ and $S_{2.1}$ exemplify the value of well-being, $S_{1.1}$ is superior to the benchmark and is an intrinsically good feature of A_1, and $S_{2.1}$ is inferior to the benchmark and is an intrinsically bad feature of A_2. I also stipulate that there is some inverse association between suffering and well-being such that $S_{1.2}$ and $S_{2.2}$ fail to instantiate the value of well-being to some degree. If the benchmark for comparison is the average of the degree to which $S_{1.2}$ and $S_{2.2}$ fail to instantiate the value of well-being, given that P suffers more in $S_{2.2}$, $S_{1.2}$ is

---

[5] Three clarifications on notation: 1) course of action A_1… has consequence S_1, courses of action $A_2…A_n$ have consequences $A_2..A_n$ respectively; 2) if A_1 has multiple consequences, I individuate these using decimal points, $S_{1.1}$, $S_{1.2}$… *Mutatis mutandis*—with the necessary changes—for $A_2…A_n$ with multiple consequences; 3) t_1, $t_2…t_n$ refer to non-identical time periods.

an intrinsically good feature of A_1, whereas $S_{2.2}$ is an intrinsically bad feature of A_2. On this analysis, A_1 ranks higher than A_2 in both dimensions of intrinsic value considered.

The second element that determines the intrinsic values of courses of action is what Wedgwood terms 'the agent's *degree of agential involvement* in bringing about each of the consequences' (334), to which there is a *causal* and an *intentional* dimension.

In respect of the causal dimension, the—perhaps familiar—idea is that there are degrees to which an agent may contribute to a state of affairs coming to pass (334-335).[6] For example, it seems plausible that in A_1, Q is more causally agentially involved in P's death through their provision of assistance to die than in A_2, in which they refuse to provide assistance to die (and presumably declines to intervene to prevent P's death by refusal of treatment). In A_1, Q ends P's life, whereas in A_2, P refuses life-prolonging treatment (and dies) following Q's refusal to assist P's death. And suppose that Q could assist P's death in one of two ways, either $A_{1.1}$: euthanasia; or $A_{1.2}$: suicide assistance.[7] It seems plausible that Q's degree of causal agential involvement is greater in $A_{1.1}$ than it is in $A_{1.2}$. For in $A_{1.1}$, Q ends P's life whereas in $A_{1.2}$, P ends their own life with Q's assistance.

In respect of the intentional dimension, Wedgwood argues that:

> Other things equal, your degree of agential involvement in bringing about a state of affairs is greater if you directly *intend* that state of affairs than if you merely *foresee* that that state of affairs will result from your action (2009, 335).[8]

---

[6] Wedgwood characterises these degrees of causal agential involvement along the lines of *failing to prevent* and *actively causing* (2009, 334), which he suggests 'many philosophers think of as the distinction between *doing* and *allowing*). For a helpful summary and critique of these distinctions, see (Woollard 2012a, 2012b). To avoid additional complication, I shall speak in terms of *relative* degrees of causal agential involvement without further classification, since, in my view, that is all we require to make Wedgwood's view work.

[7] A further clarification on notation: if course of action A_1 is pursuable by alternative means, I individuate these using decimal points, $A_{1.1}$, $A_{1.2}$… *Mutatis mutandis* for A_2…A_n pursuable by alternative means.

[8] For discussion of intentional agential involvement and reasons for action, see Wedgwood (2011).

For example, in A_1, if we grant that Q intends the course of action that involves P's death,[9] Q's degree of agential involvement is greater than it is in A_2, assuming Q merely foresees that P might refuse life-prolonging treatment (and die).

The two elements of the intrinsic values of courses of action combine in the following way: we adjust the intrinsic value of each of the consequences of a course of action relative to the benchmark for the degree of agential involvement. If S is an intrinsically bad feature of a course of action, it is made more or less bad according to the agent's high or low degree of agential involvement respectively (336). *Mutatis mutandis*–with the necessary changes—if S is an intrinsically good feature of a course of action. For example, between the options $A_{1.1}$ and A_2, if Q intends to assist P's death by euthanasia in $A_{1.1}$, this increases the goodness of the—intended and proximate—consequences $S_{1.1}$ and $S_{1.2}$. Whereas if Q intentionally refuses to assist P's death in A_2 with the foreseen and less proximate consequence that P dies by refusal of life-prolonging treatment, this decreases the badness of the consequences $S_{2.1}$ and $S_{2.2}$. x

Reasons for action

It is now possible succinctly to state the relation between intrinsic values and reasons for action. On Wedgwood's theory, 'a reason in favour of a course of action is simply an intrinsically good feature of that course of action' (336). *Mutatis mutandis* for reasons against a course of action. As explained above, whether consequence S counts as an intrinsically good or bad feature of course of action A is determined by the degree to which the relevant consequence instantiates the relevant intrinsic value, weighted for the agent's degree of agential involvement. This gives the '*agentially-weighted value*' of S (337). Whether an agent has a reason for or against A is a function of the respective positive or negative agentially-weighted

---

[9] The formulation of Q's intention here is deliberately broad to capture cases of suicide assistance in which Q does not intend that P die: (Gavaghan and King 2016, 362-364).

value of *S* (337). In addition, Wedgwood's account of reasons for action provides a basis for establishing the relative strength of reasons:

> Other things equal, the higher the positive agentially-weighted value, the stronger this reason for *A* is; the lower the negative agentially-weighted value, the stronger the reason against *A*. (The qualification 'other things equal' is important here, because there will be many cases of *interactions* between reasons…) (337).

Having outlined how intrinsic values generate reasons for action on Wedgwood's view, I am now in a position to provide an account of the *pro tanto* moral permissibility of assisted death.

The *pro tanto* case for assisted death

In this section, I make the case for the *pro tanto* moral permissibility of assisted death at a general level. I argue that, all else being equal, it is morally permissible for an agent to assist an individual to die just when their wish to die is autonomous and assisting their death is what the agent has most reason to do, where the latter is determined by the intrinsic values of the courses of action available to the agent in the individual's regard. More formally, it is, all else being equal, morally permissible for *Q* to assist *P* to die if *P*'s decision to die is autonomous and *Q* has most reason to assist *P*'s death.

For the purposes of what follows, I take as the criterion of moral rightness that a course of action is morally permissible iff—if and only if—it accords with what an agent has *most* reason to do.[10]

---

[10] This criterion prescribes a narrow range of morally permissible possible courses of action: those with the *most* agentially-weighted value. If this is a bullet, I am happy to bite it, at least in this context. Were, for example, the criterion of moral rightness that a course of action is morally permissible iff it accords with what an agent has *some* reason to do, this might entail the moral permissibility of assisted death in cases in which there is less—but still some—reason to provide assisted suicide or euthanasia relative to other courses of action available to the agent. This does not seem plausible.

I shall argue that the key to understanding the morality of assisted death lies in the intrinsic value of well-being. I then consider how this value explains the value of death and death-related reasons for action. I then explain why involuntary killing is usually morally wrong. Finally, I account for the *pro tanto* moral permissibility of assisted death.

Well-being and assisted death

The way I see it, one intrinsic value is key to the *pro tanto* moral permissibility of assisted death: the value of well-being—or prudential value or welfare. I regard the exercise of autonomy as 'one of the elements of well-being' (Mill 2002, 131-136). I shall not take a stance here on whether autonomy has instrumental—as a means—value *and* intrinsic value. If autonomy has intrinsic value (see eg Hurka 1987), it provides reasons for action relevant to the *pro tanto* moral permissibility of assisted death in its own right. But since autonomy is one of the elements of well-being, treating it independently would likely involve some degree of double counting, even if autonomy instantiates well-being in a complex way. I suggest that it is unnecessary to introduce this complication. Either way, autonomy remains a formal and substantive element of my view respectively: assisted death, by definition, requires an autonomous wish to die; and the exercise of autonomy is one way to instantiate the intrinsic value of well-being.

I shall not defend a particular conception of well-being or welfare (see eg Parfit 1984, appendix I; Tiberius 2015). The simple idea is this: well-being is that which is *good for* an individual. Their faring well to some degree is a matter of standing in a positive relation to the prudential goods that comprise well-being, and that relation is net positive against the prudential ills that comprise ill-being (Kagan 2014).

To draw the intrinsic value of states of affairs and well-being together: faring well consists in the positive instantiation of the value of well-being—however conceived—for the

state of affairs that is an individual's life at a time—synchronically—or over time—diachronically. *Mutatis mutandis* in respect of what it means to fare poorly. States of affairs that involve faring poorly *may* include chronic or terminal illness, mental disorders including anxiety and depression, trauma, the absence of pleasurable experience, the absence of love or friendship characterised by close personal relationships etc. To be clear, my claim is not that these states of affairs *necessarily* involve ill-being. Rather, I suggest that a certain kind of negative affective attitude that possibly accompanies such states of affairs is what makes for ill-being. That is, certain states of affairs may involve *suffering*, and suffering may be one of the consequences of a state of affairs that instantiates the value of well-being to some negative degree.[11] Since I only need *P* to fare poorly or to fare well to make my argument, I shall leave open what facts precisely make for states of well- or ill-being.

The value of death and reasons for action

Many of us will fare poorly—perhaps even very poorly—during some stage of our lives. Yet it does not follow that pursuing a course of action whose consequence is death is what we have most moral reason to do in such circumstances. This is because death, relative to the alternative outcomes achievable in our regard, is often thought to be *bad for us*; it deprives us of goods such as well-being that we would otherwise (expect to) enjoy.[12] Understood in this way, the value of death is *extrinsic*; it derives from:

> The difference between the intrinsic value for [*P*] of the life [*P*] would lead if [they die at t_1] and the intrinsic value for [*P*] of the life [*P*] would lead if [they die at t_2] (Feldman 1992, 150).

Helpfully, this explanation of the value of death is compatible with Wedgwood's account of reasons for action. The deprivation account relies on the determination of the relative intrinsic value of lives. We can characterise lives as complex states of affairs that

---

[11] Of course, consistent with folk accounts of the experience of suffering, suffering itself may have consequences that in turn instantiate the value of well-being to some positive or negative degree.
[12] This is commonly known as the deprivation account of the badness of death: Feldman (1991, 1992).

instantiate intrinsic values. To determine the value of death for P, we compare the relative intrinsic values of the states of affairs comprising P's life should they die at either an earlier or a later time. To determine whether Q has a reason to *cause* P's death, we compare the agentially-weighted value of the courses of action involving P dying at either an earlier or a later time with the relevant degree of Q's involvement.

On this basis, we can show that Q has a reason not to pursue a course of action that produces P's death, if, all else being equal, death would be extrinsically bad for P. Suppose that the courses of action available to Q are A_3 and A_4. A_3 involves Q pursuing a course of action that causes P's death. The relevant consequence of A_3 is S_3: P is deprived of a life in which they fare well to some degree. A_4 involves Q letting P alone, with consequence S_4: P has a life in which they fare well to some degree. It seems clear that, all else being equal, A_3 has negative agentially-weighted value, whereas A_4 has positive agentially-weighted value. In light of the value of well-being, S_3 is an intrinsically bad feature of A_3 and S_4 is an intrinsically good feature of A_4, relative to the benchmark of comparison respectively, and adjusted for Q's degree of agential involvement. Given the relation between intrinsic values and reasons for action, Q has, all else being equal, a reason against A_3 and in favour of A_4.

Mutatis mutandis, we can show that Q has a reason to pursue a course of action that produces P's death, if, all else being equal, death would be extrinsically good for P.

Why involuntary killing is usually morally wrong

Understanding how the value of well-being, the value of death, and Wedgwood's account of reasons for action relate to each other provides the basis for explaining why courses of action that involve involuntary killing are usually *pro tanto* morally wrong. I shall argue that *actual* cases of involuntary euthanasia are *pro tanto* morally permissible, however. This discussion

illustrates the general plausibility of my view and will help us to understand the morality of assisted death.

I argue elsewhere that if autonomy is one of the elements of well-being, 'every violation of autonomy results in a *pro tanto* diminishment of welfare' (Black 2018, 311). One might interpret this as a claim that P's autonomous wish to remain alive is *wholly* constitutive of their well-being. But if this were true, there could not be such a thing as involuntary euthanasia, since killing P against their wishes could not be *good for* them. All instances of involuntary euthanasia would just be instances of involuntary killing. For there to be conceptual space for involuntary euthanasia, we should instead hold that P's wish to remain alive is *partly* constitutive of their well-being or may provide *evidence* that they fare well to some degree.

I can show that involuntary euthanasia is *pro tanto* morally permissible but that other involuntary killing is *pro tanto* morally impermissible. Consider two variants of A_3 that involve involuntary killing: $A_{3.1}$ and $A_{3.2}$. On $A_{3.1}$, Q kills P with the intention that death is good for P, against P's wish to remain alive. The relevant consequence of $A_{3.1}$ is $S_{3.1}$: P is deprived of a life in which they fare *well* to some degree—because the wish to remain alive is *sufficient* to make it that they fare well or *sufficient evidence* that they fare well. On $A_{3.2}$, Q again kills P with the intention that death is good for P, against P's wish to remain alive. But $A_{3.2}$ causes $S_{3.2}$: P is deprived of a life in which they fare *poorly* to some degree—because the wish to remain alive is *insufficient* to make it that they fare well or *insufficient evidence* that they fare well. $A_{3.1}$ is just an involuntary killing. Only $A_{3.2}$ involves involuntary euthanasia.

The alternative course of action to $A_{3.1}$ is A_4: Q lets P alone with the consequence S_4 that P fares *well* to some degree. The alternative course of action to $A_{3.1}$ is A_5: Q lets P alone with the consequence S_5 that P fares *poorly* to some degree. Against their respective benchmarks of comparison, $S_{3.1}$ instantiates the value of well-being to some negative degree

and $S_{3.2}$ instantiates the value of well-being to some positive degree. For both $S_{3.1}$ and $S_{3.2}$, Q's pursuit of $A_{3.1}$ and $A_{3.2}$ involves high, albeit different, degrees of agential involvement: in $S_{3.1}$, Q is the *most proximate* cause of P's death and Q intends to kill P, but does not intend to deprive P of a life in which P fares well to some degree; in $S_{3.2}$, Q is the *most proximate* cause of P's death, Q intends to kill P and intends to deprive P of a life in which P fares poorly to some degree. Relative to A_4, pursuing $A_{3.1}$ has negative agentially-weighted value, whereas relative to A_5, pursuing $A_{3.2}$ has positive agentially-weighted value. Put in terms of—moral—reasons for action, there are, all else being equal, *pro tanto* reasons in favour of involuntary euthanasia and against other involuntary killing. It would seem, therefore, that involuntary euthanasia is *pro tanto* morally permissible.

Even if involuntary euthanasia is *pro tanto* morally permissible, it does not follow that it is morally permissible to attempt it. The reason for this is that if P's exercise of autonomy—the wish to remain alive—is partly constitutive of P's well-being or is evidence that they fare well, this will affect the degree to which the value of well-being is instantiated relative to the benchmark of comparison. If the positive intrinsic value of $S_{3.2}$ is close to the benchmark, Q ought to have less confidence that they are in fact pursuing $A_{3.2}$ (involuntary euthanasia), as opposed to $A_{3.1}$ (other involuntary killing), whose consequence is $S_{3.1}$, which has close to the benchmark negative intrinsic value. It may be the case that Q brings about an intrinsically *good* state of affairs ($S_{3.2}$) but the converse may obtain ($S_{3.1}$). And given the high degree of Q's agential involvement in $S_{3.2}$ and $S_{3.1}$ respectively, their killing of P may be *very good* or *very bad* respectively, albeit $S_{3.1}$ is less bad than $S_{3.2}$ is good. Given this uncertainty and the irreversibility of death, I argue that an agent must be *very confident* in which course of action involving involuntary killing they would in fact pursue. Since agents should rarely have such confidence, any involuntary killing is usually *pro tanto* morally impermissible, for *epistemic*

reasons. But epistemic confidence that Q pursues involuntary euthanasia may sometimes be justified. As Jonathan Glover writes:

> It does not seem plausible to say that there is no conceivable amount of future misery that would justify killing someone against his will. If I had been a Jew in Nazi Germany, I would have considered very seriously killing myself and my family, if there was no other escape from the death camps. And, if someone in that position felt that his family did not understand what the future would feel like and so killed them against their wishes, I at least am not sure that this decision would be wrong' (1990, 82).[13]

The pro tanto moral permissibility of assisted death

I am now able to account for the *pro tanto* moral permissibility of assisted death. I stated above that it is morally permissible for an agent to assist an individual to die just when their wish to die is autonomous and assisting their death is what an agent has most reason to do. This can be explained in short order.

Death is extrinsically good for P if the comparison of the intrinsic value for P of the life they would lead if they die and the intrinsic value for P of the life they would lead if they do not die is favourable to death. The intrinsic value for P of the state of affairs that is their life is determined by the intrinsic value of well-being. Autonomy is one element of the value of well-being. The intrinsic values exemplified by courses of action generate reasons for action. The direction and strength of these reasons is a function of the agentially-weighted value of the consequences of a course of action: the degree to which each consequence instantiates a relevant intrinsic value, adjusted for the agent's degree of agential involvement. An act is morally permissible if and only if it accords with what an agent has most reason to do.

Putting all this together, if among the courses of action available to Q, assisted death for P has the highest (positive) agentially-weighted value, it is what Q has most reason to do:

---

[13] Glover is speaking of the trade-off between autonomy and well-being, each of which have intrinsic value on his view. In my view, his point holds for the threshold at which we can be confident that a course of action involves involuntary euthanasia, as opposed to other involuntary killing.

assisted death is *pro tanto* morally permissible. For example, in the explanation of A_1 and A_2 above, in which the available courses of action were Q assisting P to die and Q refusing to assist P to die respectively, assisted death is *pro tanto* morally permissible, because of the way in which pursuing A_1 with the said consequences and to said degree of agential involvement most instantiates the value of well-being.

In what remains, I shall briefly treat three further relevant issues: the weight of P's wish to die, and whether there are moral differences between voluntary euthanasia and assisted suicide, and assisted death and refusal of treatment respectively.

How much credence should we give the wish to die in determining whether assisted death would, among the available courses of action, be what Q has most reason to do? Consistent with what I argue above, I take it that P's autonomous wish to die is *partly* constitutive of their well-being or may provide *evidence* that they fare poorly to some degree. It may be that P's wish to die and the other elements of their well-being align to a great degree such that no other available course of action exemplifies the value of well-being more than assisted death. In such cases—the plight of Tony Nicklinson is perhaps an apt example— assisted death may be uncontroversially morally permissible. Significantly more difficult are the cases in which P wishes to die but other elements of her well-being provide reasons against assisted death. In such circumstances, Q risks a course of action in which they assist P to die without that course of action instantiating the value of well-being more than any other available course of action.

In the event of tension between the wish to die and the other elements of well-being, I submit that the argument I made in respect of involuntary euthanasia applies, *mutatis mutandis*. If the elements of well-being—including the wish to die—conflict, any reason that Q has to assist P's death will be weaker, because the degree to which the consequences of assisting P to

die exemplify the value of well-being will be closer to the benchmark. If this is true, Q ought to have less epistemic confidence that they are in fact pursuing a morally permissible instance of assisted death. Because death is irreversible, in cases in which the elements of well-being are in tension, I argue that assisting P to die is *pro tanto* morally impermissible, and this is for *epistemic reasons*.[14]

Voluntary euthanasia involves a greater degree of causal agential involvement than assisted suicide. It seems that, all else being equal, an agent has a stronger reason to provide voluntary euthanasia than to participate in assisted suicide. The all else being equal qualification is important here. Assuming both voluntary euthanasia and assisted suicide are available, Q only has a stronger reason to provide voluntary euthanasia if it is the course of action P prefers. Otherwise, it is plausible that providing voluntary euthanasia would exemplify the value of well-being to a lesser degree than providing assisted suicide because in so doing P's autonomous wishes are fulfilled to a lesser degree.

In a similar vein, in A_1 and A_2 above, Q has more reason to assist P's death than she has reason to refuse P's death with the foreseen consequence that P refuses treatment (and dies). This example shows that assisted death may be *pro tanto* morally *preferable* to other courses of action whose consequence is P's death. It is important to note that this is not a general conclusion, but arises from the fact that in A_2, P's wish for assistance to goes unfulfilled.

Conclusion

In this chapter I advanced a *pro tanto* case for the moral permissibility of assisted death. My principal concern was the *all else being equal* moral evaluation of the dimension of assisted

---

[14] To reinforce this conclusion, I would note that when an agent refuses to assist an individual in their projects, the agent may have a low degree of agential involvement in the conseqeunces that ensue. We can see this in A_2 above.

death that comprises one individual, who wishes to die with assistance, and another, who either grants or refuses to grant the wish and who provides assistance or does not provide assistance respectively.

I argued that, all else being equal, it is morally permissible for an agent to assist an individual to die just when their wish to die is autonomous and assisting their death is what an agent has most reason to do, where the latter is determined by the intrinsic values of the courses of action available to the agent in the individual's regard. I relied on Ralph Wedgwood's (2009) account of how intrinsic values—specifically those instantiated by states of affairs—generate reasons for action, that is, how such values explain what an agent *ought* to do (for a state of affairs, at a time).

If we are to have *death rights* in the style of this volume, we require a *pro tanto* account of the moral permissibility of assisted death that provides a normative foundation for discussion of whether we have (assisted) death rights, and that partly structures the institutionalisation of assisted death (rights) in law or policy. My aim in this chapter was to advance a plausible moral theory of the permissibility of assisted death grounded in reasons for action. It is clear, however, that an account such as mine is not *sufficient* for us having a subset of death rights pertaining to assisted death. In order to do this, we must attend to those factors that are relevant to the *all things considered* morality of assisted death.

The relevant all things considered factors may include normative concerns, such as the alleged logical slippery slope from assisted death to non-voluntary euthanasia or assisted death on demand (Keown 2002; cf Lillehammer 2002), as well as mixed empirical and normative concerns, such as the 'expressivist' objection (see eg Hansard, HL Deb, 'Assisted Dying Bill [HL]' 2014 vol 754 col 809 (Baroness Campbell)), the protection of 'the vulnerable' from coercion or undue influence or ill-considered decisions (Pabst Battin et al. 2007), and the

empirical slippery slope (Lewis 2007a, 188; 2007b). We need to determine whether any of these concerns would materialise as consequences of individual instances of assisted death or of permitting assisted death in law or policy, and to determine the intrinsic value of any consequences that would obtain. Institutional design will also be highly relevant to whether assisted death is all things considered morally permissible.

I shall conclude the chapter with two general observations on the role of empirical factors and on institutionalisation as they both relate to death rights. The relevant empirical facts that influence whether assisted death is all things considered morally permissible may vary in time and space. Assisted death may not be all things considered *universally* morally permissible. However, it may be all things considered morally permissible in various places, at various times. It seems plausible, therefore, that possession and nature of death rights may be a local, rather than a universal, matter. In respect of institutionalisation, I take it that an institutional regime for assisted death that fails to approximate to a plausible moral view (and address its political justification) or attend to relevant other-regarding matters would be deficient in these senses. But the detail of institutionalising assisted death—*how* it is permitted—seems a jurisdiction-specific question—even if there may exist a global supply of possible legal ingredients (Frankenberg 2010). Death rights of the same nature may take different legal or policy forms.

Acknowledgements

References

Black, Isra. 2018. "Refusing Life-Prolonging Medical Treatment and the ECHR." *Oxford Journal of Legal Studies* 38 (2): 299-327.

Black, Isra. 2020. "Patients, Physicians and Law at the End of Life in England and Wales." In *End of Life Choices for Cancer Patients*, edited by Ruth E Board, Michael I Bennett, Penney Lewis, John Wagstaff and Peter Selby. Oxford: EBN Health.

Black, Isra. forthcoming. "Novel beings and assisted nonexistence " *Cambridge Quarterly of Healthcare Ethics*.

Coggon, John. 2010. "Assisted Dying and the Context of Debate: "Medical Law" versus "End-of-life" Law." *Medical Law Review* 18 (4): 541-563.

Daniels, Norman. 1979. "Wide Reflective Equilibrium and Theory Acceptance in Ethics." *The Journal of Philosophy* 76 (5): 256-282.

Feldman, Fred. 1991. "Some Puzzles About the Evil of Death." *The Philosophical Review* 100 (2): 205-227.

Feldman, Fred. 1992. *Confrontations with the Reaper: A Philosophical Study of Nature and Value of Death*. New York: OUP.

Finnis, John. 1995. "A philosophical case against euthanasia." In *Euthanasia examined: ethical, clinical, and legal perspectives*, edited by John Keown, xv, 340 p. Cambridge ; New York: CUP.

Frankenberg, Günter. 2010. "Constitutional transfer: The IKEA theory revisited." *International Journal of Constitutional Law* 8 (3): 563.

Gavaghan, Colin, and Mike King. 2016. "Can facilitated aid in dying be permitted by 'double effect'? Some reflections from a recent New Zealand case." *J Med Ethics* 42 (6): 361-366.

Glover, Jonathan. 1990. *Causing Death and Saving Lives: The Moral Problems of Abortion, Infanticide, Suicide, Euthanasia, Capital Punishment, War and Other Life-or-death Choices*. New ed edn: Penguin.

Harris, John. 1995. "The philosophical case against the philosophical case against euthanasia." In *Euthanasia examined: ethical, clinical, and legal perspectives*, edited by John Keown, xv, 340 p. Cambridge: CUP.

Hurka, Thomas. 1987. "Why Value Autonomy?" *Social Theory and Practice* 13 (3): 361.

Kagan, Shelly. 2014. "An Introduction to Ill-Being." In *Oxford Studies in Normative Ethics Volume 4*, edited by Mark Timmons. Oxford: Oxford University Press.

Keown, John. 2002. *Euthanasia, ethics, and public policy: an argument against legislation*. Cambridge: CUP.

Lewis, Penney. 2007a. *Assisted dying and legal change*. Oxford: OUP.

Lewis, Penney. 2007b. "The Empirical Slippery Slope from Voluntary to Non-Voluntary Euthanasia." *Journal of Law, Medicine & Ethics* 35 (1): 197.

Lillehammer, Hallvard. 2002. "Voluntary Euthanasia and the Logical Slippery Slope Argument." *Cambridge Law Journal* 61 (3): 545.

McIntyre, Alison. 2001. "Doing Away with Double Effect." *Ethics* 2 (January): 219-255.

Mill, John Stuart. 2002. *Utilitarianism and On Liberty: including Mill's essay on Bentham' and selections from the writings of Jeremy Bentham and John Austin*. 2nd edn. Oxford: Blackwell.

Pabst Battin, Margaret, Agnes van der Heide, Linda Ganzini, Gerrit van der Wal, and Bregje D Onwuteaka-Philipsen. 2007. "Legal physician-assisted dying in Oregon and the Netherlands: evidence concerning the impact on patients in "vulnerable" groups." *Journal of Medical Ethics* 33 (10): 591.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

Raz, Joseph. 1999. *Engaging reason: on the theory of value and action*. Oxford: Oxford University Press.

Scanlon, Thomas. 2008. *Moral dimensions: permissibility, meaning, blame*. Belknap.

Tiberius, Valerie. 2015. "Prudential Value." In *The Oxford handbook of value theory*, edited by Iwao Hirose and Jonas Olson. OUP.

Wedgwood, Ralph. 2009. "Intrinsic Values and Reasons for Action." *Philosophical Issues* 19: 321-342.

Wedgwood, Ralph. 2011. "Defending Double Effect." *Ratio* XXIV (4): 384-401.

Woollard, Fiona. 2012a. "The Doctrine of Doing and Allowing I: Analysis of the Doing/Allowing Distinction." *Philosophy Compass* 7 (7): 448-458.

Woollard, Fiona. 2012b. "The Doctrine of Doing and Allowing II: The Moral Relevance of the Doing/Allowing Distinction." *Philosophy Compass* 7 (7): 459-469.