2015

Two Kinds of Ends in Themselves in Kant's Moral Theory

David Hakim
Western University, davidhakim266@gmail.com

Two Kinds of Ends in Themselves in Kant's Moral Theory

Abstract

Immanuel Kant argues that rational beings are bound by an unconditional moral requirement to treat humanity always as an end and never as mere means. Kant derives this requirement from the principle that humanity is an end in itself. The purpose of my essay is to provide an interpretation of Kant's concept of an end in itself that is consistent with the other features of his moral theory and that does not have morally repugnant consequences.

To be consistent, Kant must identify a good will with an end in itself. I provide two independent arguments to demonstrate that this follows from the type of value Kant uniquely attributes to a good will and an end in itself. I also rely on Kant's teleological claims about the faculty of reason to explain how a good will can be an end in itself.

To be morally acceptable, a good will cannot be the only end in itself. If Kant identifies only a good will with an end in itself, then by "humanity" he refers only to beings that have a good will. This would mean that beings lacking a good will fall outside of the scope of the unconditional moral requirement.

I argue that Kant avoids this consequence by invoking a distinction between two kinds of ends in themselves. This distinction is suggested by a parallel between his argument regarding freedom and his derivation of the unconditional moral requirement. By attending to this parallel and the way he uses the term "humanity" in various works, I argue that Kant takes all rational beings, even if they lack a good will, to be a kind of end in itself and to deserve the fullest sort of moral consideration.

Keywords: Kant, good will, humanity, end

I – Introduction

This essay has two main goals. The first is to demonstrate that Kant identifies a good will with an end in itself. The second is to defend this interpretation (the good will interpretation) from the objection that it has morally repugnant consequences. I begin, in Section II, with a brief overview of Kant's moral theory and the formula of humanity as an end in itself (henceforth, FHE). In Sections III and IV, I put forward two independent arguments based on Kant's value claims to demonstrate that he identifies a good will with an end in itself. In Section V, I respond

to a potential objection concerning whether a good will could play the role of an end in itself in Kant's moral theory. In Section VI, I set out the main objection to the good will interpretation, namely that it has the consequence of excluding some human beings from receiving the fullest sort of moral consideration. In Section VII, I draw a parallel between Kant's argument regarding freedom and his derivation of the FHE to demonstrate that he makes a distinction between two kinds of ends in themselves. In Section VIII, I make use of this distinction to overcome the objection reviewed in Section VI.

II – The Formula of Humanity as an End in Itself

According to Kant, moral principles must be objective, meaning they must apply to all rational beings without exception (G 4:389)[1]. To satisfy this criterion, moral principles must be based on reason alone. This is because reason abstracts from everything that does not hold for all rational beings, such as desires or inclinations and the features of human nature. Human beings do not necessarily determine their will, or act, according to moral principles because their will is subject to the influence of reason as well as inclinations that can be contrary to reason. Kant claims that because the human will is imperfect in this way, it is subject to a formula called a "categorical imperative" which presents moral principles to it as absolutely necessary, or unconditional, requirements (G 4:413-4). There is only one categorical imperative but Kant provides several versions of it to represent the different features that a moral principle must have. The second version, the FHE, represents the end that a moral principle must have (G 4:436).

Kant defines an end as "that which serves the will as the objective ground of its self-determination" (G 4:427). This is commonly understood to mean that an end is something for the sake of which a rational being acts or refrain from acting (Korsgaard 185; Wood *Kant's Ethical Thought* 116). For example, in going to the gym, one's end might be health; one acts for the sake of one's health. In refraining from going out at night, one's end might be self-preservation; one refrains from acting for the sake of one's self-preservation[2]. Before Kant identifies the end of moral principles, he claims that it must be a specific type of end called an "end in itself" (G 4:428). An end in itself is an objective end, meaning that it can be held by all rational beings because it is determined by reason alone. Kant ultimately identifies the end of moral principles, or the end in itself, to be humanity. Thus, in acting on moral principles, one must act or refrain from acting for the sake of whatever is referred to by "humanity". The FHE represents this in the command to "*act* [*so that*] *you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as means*" (G 4:429). In the following two sections, I argue that Kant identifies the end in itself with a good will, and thus, takes "humanity" to refer to a good will.

---

[1] Kant's works are cited by abbreviation plus the Academy volume and page number. I use the following abbreviations: C2 = *Critique of Practical Reason*, C3 = *Critique of Judgment*, G = *Groundwork of the Metaphysics of Morals*, M = *The Metaphysics of Morals*, R = *Religion within Boundaries of Mere Reason*.

[2] I borrow this example from Korsgaard (185)

III – The Good Will and Kant's Notions of "Good" and "Value"

According to Kant, all rational beings have a faculty called a "will" which is determined, or moves a rational being to action, in accordance with principles (G 4:427). Kant defines a "good will" as a will that determines itself according to moral principles. He says that duty, i.e. acting morally, "is the condition of a will that is good" and that "[the principle of] an absolutely good will must be a categorical imperative" (G 4:403; G 4:444)[3]. Kant famously begins the first Section of the *Groundwork* with the claim that "it is impossible to think of anything at all in the world, or indeed even beyond it, that could be considered good without limitation except a **good will**." (G 4:393). He claims that, unlike a good will, one's happiness, character traits, abilities, and possessions are only good on the condition that one is worthy of them and does not put them to bad uses (G 4:393-4).

Later in the *Groundwork*, Kant claims that the end in itself is "something *whose existence in itself* has an absolute worth" (G 4:428). Richard Dean, one of the few friends of the good will interpretation, has argued that because a good will is the only thing that is good without limitation and because an end in itself has absolute value, they must be identical (35-42). This is the conclusion I am attempting to establish in this essay, but as Henry Allison points out, it is not entailed by these two value claims alone: "an unargued premise in Dean's account is the identification of being *valuable* with being *good*" (215) [4].

Though Allison is correct, there is significant textual evidence for taking Kant to endorse this premise. Kant appears to treat "good" and "valuable" as identical predicates, claiming, on the one hand, that a good will is "good without limitation" and "good…absolutely" and, on the other hand, that it "has absolute worth", "full worth in itself", and "infinite [worth]" (G 4:394; C2 5:128). Kant also says that a stoic philosopher was mistaken to call pain evil, because "pain did not in the least diminish the worth of his person" (C2 5:60). This quote indicates that Kant takes there to be a connection between evil and negative worth. Given that Kant defines evil as "the opposite contradicting [the good]" (R 6:22), he must take there to be a connection between good and positive worth. In light of what has just been reviewed, a valid argument for the identity of a good will and an end in itself (one that vindicates Dean's argument) can now be formulated:

(P1) The only thing that is good without limitation is a good will

[3] Some commentators take Kant to be referring to a perfect will rather than a good will by "absolutely good will" (Wood 15). A perfect will is one that necessarily acts on moral principles; it has no choice in the matter. I believe this reading is incorrect. Kant takes it to be impossible for a human being to instantiate a perfect will, but says that an "absolutely good will" is possible for human beings (G 4:437). Furthermore, when Kant introduces the term "absolutely good will" he strongly implies that he has returned to the topic of a merely good, rather than perfect, will (G 4:437). I take this to be good evidence for understanding Kant to use "absolutely good will" and good will interchangeably.

[4] In what follows, the terms "worth" and "value" will be used interchangeably. Oliver Sensen points out that in the original German texts, Kant has only one expression for these two terms: '*Werth'* ('Dignity and the Formula of Humanity' 102).

(P2) The terms "good" and "worth" are used as identical predicates or a connection holds between something's goodness and its worth
(P3) It follows from (P2) that something is good without limitation if and only if it has absolute worth
(P4) An end in itself has absolute worth
(P5) It follows from (P3) and (P4) that an end in itself is good without limitation
(C) It follows from (P1) and (P5) that a good will is an end in itself

IV – The Good Will as the Condition of Worth

If one remains sceptical of the connection between Kant's notions of "good" and "worth", or if, for the sake of argument, it is assumed that he uses the these notions to mean two entirely different things, it still follows from the "absolute worth" Kant attributes to an end in itself that he identifies it with a good will. The argument I make in this section to establish this conclusion runs as follows:

(P1) Something must instantiate a good will in order to have moral worth
(P2) An end in itself has a moral worth
(C) It follows from (P1) and (P2) that an end in itself must instantiate a good will

Kant does not claim that all kinds of worth depend on having some relation to a good will. For example, he takes certain things to have worth just in virtue of being desired or being useful for one's purposes: "all objects of the inclination have only a conditional worth… [and *things*] have only a relative worth, as means" (G 4:428). Although conditional or relative worth can be independent of a good will, Kant identifies a different type of worth which always presupposes, or depends for its existence, on a good will. This is expressed by Kant's claims that "nothing can have a worth other than that which is [determined by moral principles]", "all worthiness depends on moral conduct", "certainty of a disposition in accord with [moral principles] is the first condition of any worth", and "the concept of a will that is…good apart from any further purpose…constitutes the condition of all [worth]" (G 4:436; C2 5:130; C2: 5:73; G 4:397). Kant does not qualify "worth" in these quotes with expressions like "conditional" or "relative", indicating that he is talking about a distinct and moral kind of worth. These quotes also explicitly make the point that, in contrast to the former kind of worth, the latter kind depends on a good will. I contend that what Kant is expressing in these quotes is that something must instantiate a good will in order to have moral worth. This is indicated by his claim that an action must be done from duty (which is the condition of a good will) in order to have moral worth: "action first has its genuine moral worth" when it is done "simply from duty" (4:398-9). This is also indicated by Kant's claim that an individual must instantiate a good will in order to have moral worth. Kant says that when one acts from duty "it is just then that the worth of character comes out, which is moral and incomparably the highest" and that "the only thing which can give man's existence an absolute value…consists in what he does, how and on what principles he acts… in other words, I mean a good will" (G 4:438; C3 443).

In light of the preceding discussion, (P1), the premise that something must instantiate a good will in order to have moral worth, can be properly ascribed to Kant. (P2), the premise that an end in itself has moral worth, also follows from Kant's claim that it has "absolute worth" which he directly contrasts with the kind of worth that is "conditional and therefore contingent" (G 4:428). From these two premises, the conclusion follows that an end in itself must instantiate a good will. This makes perfect sense on the 'good will interpretation' because if an end in itself *is* a good will, then of course it instantiates a good will.

V – Potential Objection: Can the Good Will be an End in Itself?

Having set out my arguments for interpreting Kant to identify a good will with an end in itself, I will now turn to the question of what this identity signifies. In accordance with the explanation of an end in itself in Section II, the identity must signify that a good will is something for the sake of which all rational beings act or refrain from acting. This is *prima facie* problematic, because there are presumably some rational human beings that do not have a concept of a good will. It might seem that one must have a concept of a good will in order to know how to, and in order to be able to, act or refrain from acting for its sake. This point gives rise to a potential objection against the good will interpretation:

(P1) A rational being needs a concept of a good will to act or refrain from acting for its sake
(P2) Some rational beings lack this concept and therefore cannot act or refrain from acting for its sake
(P3) Following from (P1) and (P2) and the definition of an objective end, a good will cannot be an objective end
(P4) An end in itself must be an objective end
(C) Following from (P3) and (P4) a good will cannot be an end in itself

Kant's account of the ultimate purpose of reason constitutes an indirect response to this objection by refuting (P1). Kant makes a teleological assumption that nature has assigned all living things with instruments or faculties that are "most appropriate…and best adapted" to a certain end (G 4:395). He also claims that nature has assigned reason to us "as a practical faculty"; one that has influence over our actions (G 4:395). In accordance with Kant's teleological assumption, reason could not have been given to us to influence our actions for the end of happiness. Kant claims that instinct, rather than reason, is the most appropriate faculty for this end and he notes that people are often made less happy when they guide their behaviour by reason (G 4:396). Kant considers only one other candidate for the end of reason and concludes that reason was assigned to us to influence our actions so that we become moral rather than happy (G 4: 396). This means that the end of reason, or what all rational actions aim at, is the achievement of a good will: "the true vocation of reason must be to produce a will that is good" (G 4:396). Returning to the potential objection, a rational being does not need a concept of a good will in order to act or refrain from acting for its sake. According to Kant, when one acts from reason, this happens automatically, in accordance with the designs of nature.

VI – Main Objection: The Charge of Moral Repugnance

As mentioned in Section II, Kant explicitly identifies an end in itself with humanity. He represents this in the FHE, the unconditional command to treat whatever is referred to by "humanity" always as an end and never as mere means. From the interpretation that has been set out thus far, it follows that Kant uses the term "humanity" to refer to a good will. This would mean that, on this interpretation, the FHE doesn't command treating all human beings always as an end and never as mere means, but only the ones that instantiate a good will. Kant does not appear to believe that all human beings have a good will. He claims that "one could cite no safe examples of the disposition to act from pure duty" (G: 4:406) and suggests that at least some "human beings…are in a state of moral…depravity" (M 6: 469). Thus, at least some, and potentially all, human beings do not fall under the scope of the FHE. This morally repugnant consequence has been noted by Richard Dean: "if the good will reading is right, it looks as though we may not have [duties] towards many actual humans, since many humans lack a good will" (92). Henry Allison also articulates this concern:

> The similarities between Kant's accounts of the concept of a good will and of an end in itself… is often viewed as a source of embarrassment, since it seems to commit Kant to the implausible view that we only have duties to those with a good will (214).

VII – Two Kinds of Ends in Themselves

For now, I will set aside the objection reviewed in the previous section. Before I can adequately respond to it in defense of the good will interpretation, I must demonstrate that Kant distinguishes between two kinds of ends in themselves. I contend that Kant recognizes both an 'actual' end in itself and a 'practical' end in itself. Only a good will qualifies as an actual end in itself. It is the only thing that has absolute worth and all non-conditional, non-relative worth depends on it. It also constitutes the end of all rational actions in accordance with the ultimate purpose of the faculty of reason. By contrast, all rational beings, regardless of whether or not they use their reason to act on moral principles, qualify as practical ends in themselves. Practical ends in themselves do not necessarily have absolute worth, but, as rational beings, they are morally required to treat others and to be treated *as if* they did[5]. My argument for attributing this distinction to Kant hinges on the similarities between his argument regarding freedom and his derivation of the FHE. I will review each in turn.

Kant acknowledges that he "cannot prove…freedom as something actual" (G 4:448). Nevertheless, he claims that he can attribute a kind of freedom to all rational beings by proving

[5] Some commentators understand Kant's claims regarding the "dignity" of all rational beings to mean that they all have absolute worth, even if they do not instantiate a good will (Wood *Kant's Ethical Thought* 135). In his paper, "Kant's Conception of Human Dignity", Oliver Sensen provides compelling reasons to reject this understanding. He claims that Kant employs a Stoic conception of dignity which concerns something's prerogative over something else. Sensen catalogues all 111 uses of the term "dignity" in Kant's published writings to argue that Kant did not conceive of dignity as a non-relational value property.

that they all necessarily act under the idea of freedom. He states, "every being that cannot act otherwise than *under the idea of freedom* is precisely for this reason actually free in a practical respect" (G 4:448). For Kant, a being acts under the idea of freedom if it represents itself as free, regardless of whether or not it really is free. So, to establish that all rational beings are free in a practical respect, Kant must establish that all rational beings necessarily represent themselves as free. He does this in two steps. In the first step, he claims that an individual rational being must represent itself as free. This is because it necessarily attributes the cause of its judgments to its reason, rather than to alien influences. Kant explains that a rational being "cannot possibly think [that its] reason… would receive steering from elsewhere in regard to its judgments" and because of this, it must "be regarded by itself as free" (G 4:448). In the second step, Kant establishes that a rational being must represent every other rational being as free. Because a rational being represents itself as free in virtue of its reason, "from precisely the same ground [it has] to attribute to every being endowed with reason…this quality, to determine itself to action under the idea of its own freedom" (G 4:449). Demonstrating that all rational beings necessarily represent themselves as free, i.e. that they necessarily act under the idea of their own freedom, leads Kant to conclude that they are all practically free. "The will of a rational being can be a will of its own only under the idea of freedom and must therefore with a practical aim be attributed to all rational beings" (G 4:448).

In order to derive the FHE, Kant employs exactly the same strategy that he uses in his argument regarding freedom. He first considers how a rational being necessarily represents itself, and then argues that it is committed to representing all other rational beings in the same way:

> The human being necessarily represents his own existence in this way [as an end in itself]; so far it is thus a *subjective* principle of human actions. But every other rational being also represents his existence in this way consequent on just the same rational ground that also holds for me; thus it is at the same time an *objective* principle, from which…it must be possible to derive all laws of the will. The practical imperative will therefore be the following: *So act that you use humanity…always at the same time as an end, never merely as a means* (G 4:429)

Kant's argument regarding freedom established that all rational beings are practically rather than actually free in virtue of necessarily representing themselves as free. Given the parallel between this argument and the derivation of the FHE, the latter must only establish that all rational beings are practically, rather than actually, ends in themselves in virtue of necessarily representing themselves as ends in themselves.

VIII – Overcoming the Objection

The previous section argued that Kant makes a distinction between an actual end in itself, identified with a good will, and a practical end in itself, identified with all rational beings. It must now be considered whether Kant's use of the term "humanity" in the FHE refers to just one or both kinds of ends in themselves. At times, Kant does appear to use "humanity" to refer only to beings that have reason. For example, he claims that the predisposition to humanity "is rooted

in…reason" and he claims that the capacity to set an end by reason, possessed by all rational beings, "is what characterizes humanity" (R 6:28; M 6:392). However, Kant also uses "humanity" interchangeably with the term "personality", which involves, not just the possession of reason, but the use of reason to act on moral principles (R: 627). Kant calls personality "the idea of humanity considered wholly intellectually" (R 6:28). Christine Korsgaard notes that "in [Kant's] *Critique of Pure Reason*, humanity in one's own person and personality are spoken of as if they were the same thing" (190). Kant also refers to both "*humanity as an end in itself*" and "*personality* [as an] *end in itself*" (G 4:430; C2 587). Given that "humanity" for Kant, can refer to beings that possess a faculty reason and beings that use reason to act morally, thus instantiating a good will, "humanity" can refer to both practical and actual ends in themselves.

If "humanity" in the FHE does, in fact, refer to both practical and actual ends in themselves, then the FHE does not exclude anyone that fails to instantiate a good will from the fullest sort of moral consideration. Considering the parallel between the derivation of the FHE and the argument for freedom once more indicates that this is what Kant intended. He claims that the moral principles that "would obligate a being that is actually free are still valid for a being that cannot act otherwise than as under the idea of its own freedom" (G 4:448). By analogy, the same moral principles that obligate a being that is an actual end in itself would be equally valid for a being that is a practical end in itself.

A response to the objection reviewed in Section VI can now be put forward. If Kant took "humanity" to refer only to actual ends in themselves, then the beings that do not instantiate a good will, including at least some humans, would be denied proper moral treatment. However, Kant distinguishes between two kinds of ends in themselves and uses "humanity" in the FHE to refer to both kinds. Thus, all beings that possess a faculty of reason, and not just the beings that instantiate a good will, fall under the scope of the command to treat others and to be treated morally. Thus, interpreting Kant to identify a good will with an end in itself does not have the morally repugnant consequence it has been charged with.

IX – Conclusion

The fact that a good will is the *only* thing with a certain value, in conjunction with the fact that an end in itself has that value, entails that a good will is an end in itself. Alternatively, the fact that something must instantiate a good will to have a certain value, in conjunction with the fact that an end in itself has that value, entails that an end in itself must instantiate a good will. Because nature has ensured that our faculty of reason influences us to act morally, a good will can be something for the sake of which all rational beings act or refrain from acting, thus serving the role of an end in itself. The purpose of this essay has not been to provide a philosophical defense of any of these claims but to defend an interpretation of Kant that attributes these claims to him. Much of the secondary literature on Kant strays from his texts or glosses over these claims to provide an interpretation of his views regarding a good will, an end

in itself, and the FHE[6]. The commentators that attempt to provide a consistent and natural reading of his texts provide an interpretation of his views regarding a good will, an end in itself, and the FHE which has a morally repugnant consequence[7]. I have attempted to show, by examining the parallel between Kant's argument regarding freedom and his derivation of the FHE, and by considering how he uses the term "humanity", that a consistent and natural reading of Kant need not have this consequence. Kant holds that only a good will and the beings that instantiate it are actual ends in themselves. Kant also holds that all rational beings are practical ends in themselves just in virtue of necessarily representing themselves to be ends in themselves. By recognizing that Kant uses "humanity" in the FHE to refer to both kinds of ends in themselves, one can, at the same time, accept the good will interpretation and the fact that, for Kant, all rational beings deserve full moral consideration.

[6] The most prominent interpretations ascribe a view to Kant that is inconsistent with his claims about the good will. Korsgaard, Wood, and Allison identify all rational beings that have the mere capacity to set ends or to act on moral principles (even if they do not instantiate a good will) to be ends in themselves (see Works Cited page below).

[7] I have in mind Richard Dean, who defends the good will interpretation and bites the bullet that the FHE denies full moral consideration to beings that do not instantiate a good will (91-107).

Works Cited

Allison, Henry. *Kant's Groundwork for the Metaphysics of Morals: A Commentary*. Oxford: Oxford University Press, 2011. Print.

Dean, Richard. *The Value of Humanity in Kant's Moral Theory*. Oxford: Oxford University Press, 2006. Print.

Kant, Immanuel. *Critique of Judgment*. Trans. and Ed. Werner Pluhar. Cambridge: Hackett Publishing Company, 1987. Print.

---. *Critique of Practical Reason*. Trans. and Ed. Mary Gregor. Cambridge: Cambridge University Press, 1997. Print.

---. *Groundwork of the Metaphysics of Morals*. Trans. and Ed. Mary Gregor. Cambridge: Cambridge University Press, 1998. Print.

---. *The Metaphysics of Morals*. Trans. and Ed. Mary Gregor. Cambrige: Cambridge University Press, 1996. Print.

---. *Religion within the Boundaries of Mere Reason*. Trans. and Ed. Allen Wood and George Di Giovanni. Cambridge: Cambridge University Press, 1998. Print.

Korsgaard, Christine. "Kant's Formula of Humanity." *Kant-Studien*. 77.2 (1986): 183-202. Print.

Sensen, Oliver. "Dignity and the Formula of Humanity." *Kant's Groundwork of the Metaphysics of Morals: A Critical Guide*. Ed. Jens Timmermann. Cambridge: Cambridge University Press, 2009. 102-118. Print.

Wood, Allen, "The Good Without Limitation." *Groundwork for the Metaphysics of Morals*, Eds. Dieter Schönecker and Christoph Horn. Berlin: Walter de Gruyter, 2006. 1-27. Print.

Wood, Allen. *Kant's Ethical Thought*. Cambridge: Cambridge University Press, 1999. Print.