Kantian Forgiveness: A Shared End

Sofía Stutz

Posner Fellow, Northwestern University

August 14th, 2020

Suppose Tom is trying to clear up space in the house. He decides to get rid of what he perceives to be clutter by selling some of his wife Meg's old books without asking her. He assumes she won't notice or mind. But she does, reproaching him for selling her novels without asking her first. Meg is angry and resentful 1) because the exact novels are irreplaceable (and of great sentimental value), and most importantly 2) because he did not consult her before making the decision, disregarding her authority. If Tom refuses to apologize, is Meg still obligated to forgive him? What if he apologizes sincerely but she nevertheless refuses to forgive him? Is an apology a necessary or sufficient condition for forgiveness? Under what conditions is forgiveness a Kantian duty, if ever? To explore the answers to these questions, I will limit the scope of this essay to forgiveness as it pertains to interpersonal relationships, particularly between two individuals. I will also focus on wrongs of less moral gravity than that of atrocities such as murder, rape, and torture. It seems unreasonable to lump together "serious" and "less serious" wrongs, and impossible to adequately cover both within the bounds of this paper. For this reason, I will primarily examine the latter category.

That being said, I hold that no person—whether he commits minor or major moral offenses—is absolutely unforgivable, and that *people*, not deeds, must be the objects of our forgiveness. I argue that the Formula of Humanity calls us to not only reject the claim that some people are absolutely unforgivable, but to put forgiveness into practice ourselves, treating it as a shared end in our interpersonal relationships. Provided that the original agents involved are alive and lucid, the offender repents and apologizes sincerely, and the victim forgoes resentment, offenders and victims should consider taking up forgiveness collectively.

In section one, I will define forgiveness. In section two, I will summarize Trudy Govier's argument for the rejection of absolute unforgivability, explaining why I believe we must take her argument a step further. In section three, I will address what Kant writes about forgiveness. I will also explain how the Formula of Humanity demands our acknowledgement of humanity, and in the same vein, encourages the practice of forgiveness. I will argue that the Kantian duty to promote the happiness of others can mean taking up forgiveness as a shared end in our interpersonal relationships.

Section I: What is forgiveness?

Most recent philosophical accounts of forgiveness assert that sincere forgiveness requires the renunciation of resentment.[1] Andrea Westlund refers to this as the standard approach.[2] Pamela Hieronymi modifies the standard approach, qualifying it with the following judgments:

> (1) The act in question was wrong; it was a serious offense, worthy of moral attention. (2) The wrongdoer is a legitimate member of the moral community who can be expected not to do such things. As such, she is someone to be held responsible and she is worth being upset by. (3) You, as the one wronged, ought not to be wronged. This sort of treatment stands as an offense to your person.[3]

Hieronymi claims that resentment can be best understood as protest against an offense.[4] Furthermore, resentment affirms the offender's moral significance and that of the person wronged. She argues that an adequate account of forgiveness entails articulating how the forgiver

---

[1] Andrea C. Westlund, "Anger, Faith, and Forgiveness," *The Monist* 92, no. 4 (2009): 507.
[2] Westlund, "Anger, Faith, and Forgiveness," 507.
[3] Pamela Hieronymi, "Articulating an Uncompromising Forgiveness," *Philosophy and Phenomenological Research* 62, no. 3 (2001): 530. https://doi.org/10.1111/j.1933-1592.2001.tb00073.x.
[4] Hieronymi, "Articulating an Uncompromising Forgiveness," 530.

can abandon protest while simultaneously maintaining the judgments above.[5] I accept Hieronymi's standards for an uncompromising account of forgiveness, and will apply her judgments accordingly. Ultimately, I believe that for offenders, seeking forgiveness entails genuine repentance and apology, the latter of which is a necessary (but not sufficient) condition given that it formalizes the offender's acknowledgment that his action was morally unacceptable and communicates that sentiment to the person wronged. The wrongdoer's apology is not sufficient because forgiveness still requires the victim's input. To put it idiomatically, it takes two to tango; in order to grant forgiveness, the wrongdoer must forgo her resentment.

Section II: Extending Govier's argument

Govier holds that forgiveness and unforgivability apply to agents and not to their deeds.[6] In other words, forgiveness does not entail condoning a perpetrator's deeds nor does it absolve a perpetrator from appropriate punishment. Forgiveness applies to *people* who feel remorse as they recognize the immorality of their deeds and the pain they have caused other people. Those who have committed wrongs, refusing to acknowledge their wrongdoing toward victims and their communities, are *conditionally unforgivable*; they are unforgivable on account of their refusal to recognize their wrongdoing, transform their moral attitudes, and offer restitution to their victims.[7] Nevertheless, no person is *absolutely unforgivable*—a term that, if applied to a perpetrator, means that it "would never under any circumstances be morally appropriate for anyone to forgive him."[8] Govier explains that "in its attribution of moral fixity and failure to recognize the offender's capacity for moral change, the unforgiving attitude amounts to an attitude of

[5] Hieronymi, "Articulating an Uncompromising Forgiveness," 530.
[6] Trudy Govier, "Forgiveness and the Unforgivable," *American Philosophical Quarterly* 36, no. 1 (1999): 59.
[7] Govier, "Forgiveness and the Unforgivable," 71.
[8] Govier, "Forgiveness and the Unforgivable," 67.

disrespect for the wrongdoer as a person."[9] All people, regardless of their actions in the past, have the potential to change, and for this reason, we must reject the notion of absolute unforgivability.[10]

For serious wrongs, the stringent distinction between a person and her deeds may seem counterintuitive. There's a tautological sense in which people who murder, for instance, are murderers.[11] Nevertheless, people who have murdered are not *only* murderers; claiming that they are reduces them purely to their wrongs.[12] All human beings, no matter what they've done in the past, possess a potential for change. For the purposes of this essay, I will not venture into forgiveness related to moral atrocities or conflict between more than two individuals. Nonetheless, I believe Govier's claim about the potential for moral transformation carries weight, both in extreme cases and less serious ones. Denying the transformative capacity she refers to would be tantamount to negating human autonomy and rational will—key tenets of Kantian moral theory, which I will examine further in section three.

While I concur with Govier's claim about absolute unforgivability, her argument does not go far enough in that she does not examine one of its implications. Namely, that from the rejection of absolute unforgivability follows not an *obligation*, but a strong *normative pull* toward forgiveness. If we claim that a wrongdoer is not absolutely unforgivable, and the wrongdoer repents and seeks forgiveness (eliminating the possibility of conditional unforgivability), then the next logical step seems to be that the person wronged should forgive or at the very least consider forgiving her wrongdoer. If we limit the scope of this analysis to

---

[9] Govier,"Forgiveness and the Unforgivable," 62.
[10] Her account is not necessarily an optimistic one. It does not posit that people are *likely* to change. It only claims that all people possess the potential to change and that we have a duty to acknowledge this.
[11] Govier,"Forgiveness and the Unforgivable," 66.
[12] Ibid., 66.

conflict between original agents, physically capable of engaging in dialogue (in other words, alive and lucid), then naturally only the person wronged can forgive. If one or both of the original parties is no longer alive or lucid, then a secondary representative (defined as a family or community member close to the original victim) might initiate or continue dialogue on behalf of the original agents.

Returning to our example, you might imagine that Meg declares: "I'm not saying Tom can't be forgiven. I'm just saying *I* don't want to forgive him." Meg claims to reject the notion of absolute unforgivability but, at least in that particular moment, she does not feel inclined to forgive Tom. Her resentment is warranted; she has been disrespected and it is reasonable to allow her time to express her anger and cool off. And in order for genuine forgiveness to take place, she must not compromise the judgments Hieronymi outlines; to excuse Tom's action and disregard her own feelings would be to undermine her rational grounds for resentment.

As a manifestation of anger, Meg's statement is reasonable. However, if she relentlessly refuses to forgive Tom yet continues to stand by this claim, her statement seems problematic in the end. Refusing to forgive Tom or even refusing to *consider* forgiving Tom ultimately amounts to the same outcome as one dictated by the person who accepts the notion of absolute unforgivability. Since she is alive and lucid, she cannot defer the task of forgiveness to anyone else; she is the only person in this scenario in a position to grant Tom forgiveness. A friend might point out the discrepancy and Meg might reply: "Okay! I revoke my rejection of absolute unforgivability. In this case, Tom is absolutely unforgivable." But as Govier claims, absolute unforgivability is unacceptable. Meg cannot truly appeal to this concept to correct the inconsistency since absolute unforgivability fails to be philosophically viable.

So far we have claimed that we must reject the concept of absolute unforgivability, which I will break down in Kantian terms in section three. We have also claimed that if you refuse to forgive or consider forgiving your offender—whether you reject or claim to accept the concept of absolute unforgivability—you are implicated in a somewhat inconsistent position. How so? Since absolute unforgivability is unacceptable, the possibility of forgiveness should not be *discounted* in cases where conditional unforgivability does not apply. In such cases, only those offended are in a position to forgive. While these statements do not indicate that the person wronged *must* forgive her wrongdoer, they do suggest that perhaps the person wronged has good reason to consider forgiveness as a compelling *possibility* that aligns with the rejection of absolute unforgivability, insofar as she maintains Hieronymi's judgments.

The steps above do not explain, however, *why* the person wronged should consider taking up forgiveness as a shared end. There's a gap between acknowledging someone's humanity to affirm their transformative capacity and undertaking full-fledged forgiveness collectively. Govier argues that no person is absolutely unforgivable but does not go so far as to say that victims *must* forgive their perpetrators. To argue against the notion of absolute unforgivability, she invokes the Kantian principle that we *must* acknowledge humanity. While they are closely linked, however, forgiveness and acknowledgement of humanity are not synonymous phenomena. In section three, I will attempt to bridge the gap between the duty to acknowledge humanity, as posited by the Formula of Humanity, and forgiveness, which I believe can fall under the Kantian duty to promote the happiness of our loved ones.

Section III: Kant on Forgiveness and the Formula of Humanity

Govier insists we have a duty to affirm others' transformative capacity in virtue of their humanity. What exactly does this mean in Kantian terms? In the second section of the *Groundwork*, Kant introduces the Formula of Humanity, claiming that humans are ends in themselves.[13] Humans have *dignity*, which for Kant, denotes irreplaceability. Things that have a *price*, on the other hand, are replaceable because they have an exchange value. As Kyla Ebels-Duggan explains, "no exchange could compensate for the loss of a person."[14] All humans have equal value but this does not mean we can be replaced. People are intrinsically valuable and for that reason, must never be treated as means; Kant's practical imperative demands: "so act that you use humanity, in your own person as well as in the person of any other, always at the same time as an end, never merely as a means."[15]

The Formula of Humanity is a statement or formulation of Kant's Categorical Imperative (CI), which demands that you "act only according to that maxim through which you can at the same time will that it become a universal law."[16] In other words, act only in ways that would not result in a contradiction if everyone did it. Kant also posits the principle of autonomy—a different way to formulate the very same moral law of the CI. The principle of autonomy claims that "the will of every rational being is a universally legislating will."[17] Here Kant argues that the rational will is a law to itself; the will gives itself lawlike maxims on which to act. When we respect the law, we embody an attitude toward our own rational will and that of other human

---

[13] Immanuel Kant, *Groundwork of the Metaphysics of Morals*, trans. Mary J. Gregor (Cambridge, U.K. ; New York: Cambridge University Press, 1998), 41.
[14] Kyla Ebels-Duggan, "Dignity and Consequences," *Norton Introduction to Ethics*, 1.
[15] Kant, *Groundwork of the Metaphysics of Morals*, 41.
[16] Kant, *Groundwork of the Metaphysics of Morals*, 34.
[17] Kant, *Groundwork of the Metaphysics of Morals*, 43.

beings. Kant defines the autonomy of the will as the supreme principle of morality, and the ground of dignity in rational beings.

Respecting one another's inherent capacity as rational beings to make maxims for ourselves means we affirm each other's autonomy, and therefore, our ability to make new choices in the future. Since the Formula of Autonomy is a statement of the CI, and the Formula of Humanity is also a statement of the CI, the recognition of autonomy is also a recognition of human dignity. For this reason, a victim has a duty, in virtue of the perpetrator's autonomy, to reject dehumanizing the perpetrator. Again, this does not entail condoning or excusing her morally reprehensible deeds.

A victim must only acknowledge the perpetrator's autonomous and rational will, which in Kant's view, is a person's true self—the "intelligible essence of a person," as J. David Velleman puts it.[18] Kant describes the rational nature of humans as a self-existent end; since it already exists, it cannot be produced.[19] Velleman clarifies that "rational nature" refers to the capacity to operate on the basis of reason and to have good will (in other words, acting from duty and for the right reasons rather than acting on the basis of an action's consequence). This is what constitutes the true self. Velleman emphasizes that the rational nature does not refer so much to the intellect as to people's "core of reflective concern," i.e, their ability to appreciate the value of ends.[20]

In practice, acknowledging others' rational nature means rejecting the notion of absolute unforgivability, and recognizing a perpetrator's potential for change. In refusing to do so, the victim implies that the wrongdoer is incapable of change, and this amounts to disrespect in its

[18] J. David Velleman, "Love as a Moral Emotion," *Ethics* 109, no. 2 (1999): 344, https://doi.org/10.1086/233898.
[19] Velleman, "Love as a Moral Emotion," 357.
[20] Velleman, "Love as a Moral Emotion," 365-366.

denial of the wrongdoer's humanity and rational nature. Like Govier, I defend human transformative capacity in virtue of the autonomous, rational will at the center of Kant's moral system.

In the Doctrine of Virtue in the *Metaphysics of Morals*, Kant argues that to deny that someone who has acted badly in the past can act well in the future is to violate a duty of respect.[21] He explicitly states that it is a "duty of human beings to be forgiving" because "a human being has enough guilt of his own to be greatly in need of pardon" and because no punishment should "be inflicted out of hatred."[22] Immediately after claiming this, however, Kant warns us against distorted forgiveness:[23]

> But this [duty] must not be confused with meek toleration of wrongs … for then a human being would be throwing away his rights and letting others trample on them, and so would violate his duty to himself.[24]

Steering clear of this kind of forgiveness, which may be better described as a lack of self-respect, Kant claims forgiveness to be a duty. Hieronymi's judgments are helpful as they prohibit the distortion Kant warns against, affirming the wrongfulness of the deed in question, and the moral legitimacy of both the wrongdoer and the one wronged.

Kant depicts the process of forgiveness as one of grace and mercy. His phrase "greatly in need of pardon" evokes an image of vulnerable supplication on the part of the wrongdoer and elective power on the part of the victim. By claiming forgiveness to be a duty, Kant does not seem to equate it with obligation. Personal autonomy is key to his moral system and hence,

---

[21] Immanuel Kant, *The Metaphysics of Morals*, trans. Mary J Gregor. Revised edition.. Cambridge, United Kingdom: Cambridge University Press, 2017.
[22] David Sussman, "Kantian Forgiveness." *Kant-Studien.* 96, no. 1 (2005): 85.
[23] Sussman, "Kantian Forgiveness," 88.
[24] Ibid., 88.

forgiveness should never impinge upon the victim's freedom; it is a gift she grants freely. Nevertheless, Kant expresses a level of urgency and necessity as he describes forgiveness. We are greatly in *need* of pardon, he insists.

Bearing in mind the Categorical Imperative in all its formulations, as well as Kant's specific statements on forgiveness, the following question remains: how do we jump from the duties laid out in the CI to forgiveness? To answer this question, we must introduce a new element to the equation. The fourth example Kant outlines in the *Groundwork* exemplifies the duty to promote the happiness of others.[25] Kant claims:

> "For the ends of any person, who is an end in himself, must as far as possible be my end if that conception of an end in itself is to have its full effect on me."[26]

Christine Korsgaard explains that this is "because the full realization and acknowledgment of the fact that another is an end in itself involves viewing the end upon which the person confers value as good."[27] To acknowledge another's humanity—to treat that person as an end in itself—entails that we treat "his or her ends as objectively good, as you do your own."[28]

Ebels-Duggan applies this principle to our dynamics with our loved ones.[29] Asserting that happiness is synonymous with well-being, she argues we should aim to share our loved ones' ends. We are not required to treat one another's judgment as infallible, she explains, but we must give each other the benefit of the doubt, assuming that each other's choices are good. In other words, we must grant each other trust. While we should share in our loved ones' ends, she

---

[25] Christine M. Korsgaard, *Creating the Kingdom of Ends* (New York, NY, USA: Cambridge University Press, 1996), 127.
[26] Korsgaard, *Creating the Kingdom of Ends,* 127.
[27] Ibid., 127.
[28] Ibid., 127.
[29] Kyla Ebels-Duggan, "Against Beneficence: A Normative Account of Love," *Ethics* 119, no. 1 (2008): 142. https://doi.org/10.1086/592310.

clarifies, we should refrain from adopting ends that place significant and unreasonable demands on each other. We must set ends that include our partner's input. Doing otherwise would allow us to overstep the bounds of our own authority within the relationship. Ultimately, partners must strike a balance between choosing ends somewhat jointly while also respecting each other's individual authority.

To return to our running example, I believe Meg and Tom—and all those committed to fostering egalitarian interpersonal relationships—should approach forgiveness as a shared end. A key to all functioning intimate relationships, forgiveness is a reasonable end to strive for reciprocally. Ebels-Duggan's approach posits an egalitarian view that demands mutual respect of authority and a commitment to collective discernment toward a single set of shared ends. Under this framework, Tom's supplication would be a recognition of Meg's mutual authority. This is something she would likely want since precisely what he compromised in selling her books was his recognition of and respect for her authority in the relationship.

The shared-ends view allows for the kind of flexibility necessary to a convincing account of forgiveness; we must fulfill certain conditions if we intend to take up forgiveness as a shared end but the interpretation of Kant's stance is not restrictive in the sense that it *obligates* victims to grant their wrongdoers forgiveness, which would render it antithetical to personal freedom and in many cases require victims to alter their emotions on demand—an unreasonable expectation. The Kantian duty to promote our loved ones' ends *encourages* rather than forces us to consider taking up forgiveness as a mutual aim. The shared-ends view takes into account the reality that renouncing resentment, repenting, and forming sincere apologies isn't always immediate. The position allows space for a range of scenarios and responses to wrongdoing while simultaneously

providing a compelling reason for the people involved to collectively undertake the project of forgiveness. Namely, that forgiveness is something we should strive for if we take our commitment to healthy and egalitarian relationships seriously.

Conclusion

Applying the shared-ends view to forgiveness, perhaps it is now possible to answer some of our initial questions. I queried whether or not Meg is obligated to forgive Tom if he refuses to apologize. His unwillingness to recognize his wrongdoing, transform his moral attitude, and offer restitution to Meg in some measure would allow her to classify him as conditionally unforgivable. If instead he apologizes sincerely yet she refuses to forgive him, she disregards the normative pull of forgiveness and the cogent reasons for reciprocating his attempt to treat forgiveness as a shared end. Under no condition, however, is she *obligated* to forgive him—nor is anyone ever under such an obligation since forgiveness is a gift. Ultimately, sincere repentance and apology on the part of the offender, renunciation of resentment on the part of the victim, and both parties' capacity to engage in dialogue are the conditions under which the parties in question can undertake forgiveness together. Not only must we reject the notion of absolute unforgivability, but we should also put forgiveness into practice, treating it as a collaborative and worthwhile project.

Works Cited

Ebels‐Duggan, Kyla. “Against Beneficence: A Normative Account of Love.” Ethics 119, no. 1 (2008): 142–70. https://doi.org/10.1086/592310.

Ebels-Duggan, Kyla. “Dignity and Consequences.” Forthcoming in the *Norton Introduction to Ethics*.

Govier, Trudy. “Forgiveness and the Unforgivable (Moral Philosophy).” *American Philosophical Quarterly* 36, no. 1 (1999): 59–75.

Hieronymi, Pamela. “Articulating an Uncompromising Forgiveness.” *Philosophy and Phenomenological Research* 62, no. 3 (2001): 529–55. https://doi.org/10.1111/j.1933-1592.2001.tb00073.x.

Kant, Immanuel, and Gregor, Mary J. *Groundwork of the Metaphysics of Morals*. Cambridge, U.K. ; New York: Cambridge University Press, 1998.

Kant, Immanuel, Denis, Lara, and Gregor, Mary J. *The Metaphysics of Morals*. Revised edition.. Cambridge, United Kingdom: Cambridge University Press, 2017.

Korsgaard, Christine M. *Creating the Kingdom of Ends*. Cambridge; New York, NY, USA: Cambridge University Press, 1996.

Sussman, David. “Kantian Forgiveness.” *Kant-Studien.* 96, no. 1 (2005): 85–107.

Velleman, J. David. “Love as a Moral Emotion.” *Ethics* 109, no. 2 (1999): 338–74. https://doi.org/10.1086/233898.

Westlund, Andrea C. “Anger, Faith, and Forgiveness.” *The Monist* 92, no. 4 (2009): 507–36.