A Kantian Account of Moral Trust

Eli Benjamin Israel

November 27, 2024 Penultimate draft.
Forthcoming in Kantian Review. Please cite published version.

Abstract: In this paper, I propose a Kantian framework for moral trust—trust in another person to only act with us in morally permissible ways. First, I derive an understanding of trustworthiness from Kant's second formulation of the categorical imperative. I argue that trustworthiness embodies a moral imperative, guiding us to act in ways that are reliable and recognizable as conducive to engaging in trusting relations. However, this alone is not enough, as it doesn't provide a means to assess whether someone is truly committed to the moral law and thus morally trustworthy. Therefore, in the second part, I explore a basis for assessing their moral conduct found in a local version of the Kingdom of Ends: given an ideal or archetype of a morally perfect interpersonal relationship, an archetype of the morally trustworthy agent allows us to comparatively assess the moral disposition of fellow agents.

1. Introduction

Respect takes center stage in Kant's moral philosophy. It is a primary virtue we owe to others, requiring us to recognize the autonomy of fellow individuals, acknowledge their ends, affirm their status as moral agents, and ultimately act morally with them. However, this emphasis on respect makes morality a one-way street, focused on our universal moral obligations toward others, without addressing the conduct of others in our practical reasoning.

The consideration of other people's actions, I argue, is necessary for a comprehensive account of the practical domain, not because one's moral commitment is conditional on reciprocal expectations, but because important aspects of this domain are two-way streets and

For discussions and comments on earlier versions, I thank Brian Hutler, Colin Marshall, Hans Shenk, and Kevin Doherty, the anonymous reviewers for Kantian Review, as well as the audience at the North American Kant Society 2024 Pacific Study Group Conference.

depend on how others will act. This oversight of others' actions within practical reason is not unique to Kant but rather a general concern that applies to other moral theories as well. However, it is particularly striking to think of Kantian ethics lacking in this way, given Kant's strong emphasis on others as acting agents and the communal aspects of his practical philosophy, as we find in the second and third formulations of the categorical imperative, requiring us to acknowledge and constraint our actions in light of the agential presence of others in our practical environment, acknowledging them as ends in themselves and co-legislators of morality, respectively. Furthermore, Kant's writings include detailed discussions of the moral community, which one might argue would make us expect it to actively consider the dynamic interplay of moral conduct among the members of such a community.

Aspects of our practical reasoning can be described as "two-way streets" in the sense that, when determining our actions, in addition to internal consideration regarding our motives, we often also have to make external considerations regarding the conduct of others. These considerations become more salient in scenarios where we interact and build upon each other's actions, and so we depend on how others will act for the successful achievement of our ends, being those moral or merely prudential. In such cases, we may find ourselves having to assess whether another person will act morally so that we can plan our actions accordingly. Such consideration of others' actions, I argue, can be well introduced into our practical reasoning through the notion of interpersonal trust, and as we are concerned, in particular with their moral conduct, in this paper I develop an account of "moral trust."

I define moral trust as reliance on a co-agent's moral conduct toward us. This aligns with prevalent accounts of trust in contemporary literature which take it to be fundamentally a motive-based attitude (Baier 1986; Jones 1996; O'Neill 2002). However, the true significance of a Kantian account of moral trust is that it provides our practical reason grounds for relying

on others' actions that are rationally and universally valid, instead of relying on "contingent, subjective conditions that distinguish one rational being from another" (*KpV* 5:21). If I can morally trust another person, I can directly access what they can be trusted to do even when in unexpected situations which we haven't discussed. Therefore, moral trust provides us with confidence about our trustees' actions

To illustrate how this form of trust may come into play, think for a moment of that friend of yours whom you trust not due to personal loyalty, but because you believe they will always do the right thing. Let's say you confide to them about a moral lapse, like infidelity. If you morally trust this friend, you can't rely on them to help you hide your indiscretion from your partner. Nonetheless, you can trust them to confront you about the moral breach and urge you to make things right. They are also someone you can trust not to break a promise made to you, even if your friendship were to end, and you can confidently recommend their services to a family member because you trust that, regardless of contingent considerations and whether they know who brought them that client, they will always act morally with their customers.

Surprisingly, Kant gives no attention to this form of trust, even though he acknowledges the inherent social dimension of morality, asserting that it is a duty to employ one's moral capacities in establishing conditions for improved social interactions. In his words:

> It is a duty to oneself as well as to others not to isolate oneself (*separatistam agere*) but to use one's moral perfections in social intercourse (*officium commercii, sociabilitas*). [...] Not exactly in order to promote as the end what is best for the world but only to cultivate what leads indirectly to this end: to cultivate a disposition of reciprocity — agreeableness, tolerance, mutual love and respect (affability and

> propriety, *humanitas aesthetica et decorum*) and so to associate the graces with virtue To bring this about is itself a duty of virtue. (*MS* 6:473)

Moreover, the pivotal role of social interactions in Kant's moral framework finds its culmination in his notion of the kingdom of ends, as "a systematic union of various rational beings through common laws" (*G* 4:434). This ideal represents a social order fully guided by the moral law, serving perhaps as a fourth formulation of the categorical imperative.

I address this lacuna in this paper, suggesting moral trust as a complementary facet of Kant's moral philosophy. I argue that not only that it is consistent with the resources Kant provides in his moral writings, but it also aligns with and helps to realize his envisioned objectives for an interpersonal, socially-directed, two-way street morality.

My argument is divided into two parts, addressing respectively the normative and epistemic aspects that arise when reconstructing such an account. First, I establish an account of moral trustworthiness derived from Kant's second formulation of the Categorical Imperative, also known as the Formula of Humanity (*FH*). I argue that *FH* embodies a duty of acting in ways that are reliable and recognizable as conducive to engaging in trusting relations, thus addressing the normative question of how one ought to behave to be morally trustworthy. However, this alone is insufficient for a comprehensive account of moral trust, as it does not address the challenge of determining whether another person is genuinely committed to acting morally. Therefore, in the second part of my argument, I explore a method of assessing the trustworthiness of others through an ideal or archetype of a morally trustworthy agent, found in a local version of the Kingdom of Ends. This addresses the epistemic puzzle of how we can know or reasonably judge whether someone should be morally trusted.

2. Ground Clearing

Trust is a complex concept that can be associated with a variety of phenomena, whether they are feelings or attitudes, cognitive or conative, voluntary or not, directed towards objects, persons, or collective entities, and so forth. In this paper, I discuss one particular conception of trust—moral trust. While this focus may not encompass the diverse usages of trust, it is tailored to illuminate its distinct contribution to our practical reasoning within a Kantian framework. Nevertheless, I believe the idea of "moral trust" is consistent with the main trends in recent philosophical literature.

In the contemporary literature, it's widely accepted that trust is more than mere reliance (Marušić 2017; McLeod 2021).[2] Reliance is something I can attribute to objects, such as when relying on my alarm clock to wake me up in the morning; or to people *qua* object, that is, when I predict that they will act in certain ways without requiring their active involvement in that prediction, as when I rely on the fact that my wife usually arrives home before me because I forgot my keys. These cases are characterized by an expectation or confidence I hold in things going in certain ways, without requiring any commitment or responsibility on their part. Cases of reliance, however, lack the vulnerability to another person's will that we find in trust. They aren't actively involved in our expectations of them, and therefore, we don't consider them responsible and accountable for not acting according to our expectations. If my alarm clock's battery dies or my wife has an important meeting I didn't know about, it wouldn't be appropriate to hold them responsible for me not waking up or waiting outside, respectively.

[2] Among those contemporary views, some of them have particularly strong Kantian leanings, as can be seen in Baier (1986), O'Neill (2002), and other motive-based theories, which assume that what is added to reliance is goodwill on the part of the trustee towards the trustor.

Trusting someone involves a kind of confidence in the trustee that goes beyond a mere prediction of what they will do. We acknowledge that it is up to them to fulfill our trust and act as we expect from them. In addition, we consider them to owe us the relevant action. This underscores the vulnerability inherent in trust, where we invest not only an expectation of a certain outcome but also a dependence on the trustee's commitment and accountability.

Thus, trust implies a duty, a normative commitment on the part of the trustee. This duty may derive from personal relationships, moral commitment, or other principles a person is committed to within their interpersonal and social systems. The idea the trustee holds a commitment to act in certain ways is likely what grants trust its characteristic dynamics, including the possibility of holding the trustee accountable for their commitment, and the potential for the trustor to experience feelings of betrayal or personal disappointment if a violation of trust occurs (Baier 1986, 235; Holton 1994, 4).[3] Moral trust, in particular, implies a commitment or respect for the moral law, for which the trustee is expected to act morally, and the trustor may hold them accountable the trustee accountable if they fail to fulfill their commitment, at least regarding failures with practical relevance to the trustor, constituting a moral failure towards them.

Kant doesn't provide a sustained treatment of trust in his moral writings. Scattered remarks involving trust appear in his later works, but these do not explicitly address trust as an interpersonal, and potentially morally-based phenomenon. Previous discussions of trust in Kantian scholarship go in the same ways, exploring trust in specific contexts that don't

[3] This normative stand in moral trust is of moral nature, of course, but it might actually be moral also in more general accounts of trust as well. It is true that, in a sense, one may trust another to act contrary to what is morally permissible, and in this context, acting morally would be a violation of that trust. And still, there seems to be something fuzzy about such a scenario that doesn't seem to conform to the idea of trust. At least from a Kantian perspective, it seems to require someone to act out of heteronomous motives, instead of autonomous ones that ultimately require their compliance with the moral law. Because of this, at least upon initial consideration, morally condemning attitudes associated with breaches of trust, such as resentment, appear to be unjustified in such cases.

ground the term in the moral law, such as within Kant's theory of international relations (Schröder 2010); or religious faith (Sussman 2001). These context-specific accounts are not directly relevant to our analysis.

An exception can be found in Longworth (2017), who develops a Kantian framework for interpersonal trust, exploring whether we may trust someone without evidential grounds for their trustworthiness, and ultimately argues that it's reasonable, from a pragmatic standpoint, to do so. Assuming that interpersonal trust means that one holds true that the other person can be relied upon, Longworth draws on Kant's conception of moral *Glaube*, in which one is rationally justified in holding something as true, e.g., the existence of God or the immortality of the soul, on practical rather than evidential grounds. Specifically, what justifies belief in these propositions is the coherence of practical reason and the pursuit of moral ends, particularly the realization of the highest good. Then, he applies this form of rational justification to the discussion on trust, concluding that it's reasonable, from a Kantian perspective, to hold true that someone is trustworthy without some form of assessment of their trustworthiness. He argues that one may place their trust in others merely in light of some practical end that requires their cooperation and the lack of "too much evidence that they are untrustworthy or unreliable" (2017, 268). In other words, Longworth contends that we are justified in trusting others solely based on our practical interest to do so, as long as our positive presumption about them is not contradicted by significant countervailing evidence. Thus, he grounds trust solely on non-epistemic grounds, that is, on why it would be valuable or desirable if we could trust that person.

Now, this view strikes me as extremely counterintuitive. One may indeed have non-epistemic practical grounds for trust — i.e., understand that trusting another can help them to achieve their ends. However, these alone are insufficient for trust to be justifiably established, and so, rational. When we trust, we care about whether it is true that our trustees will act in

trustworthy ways because the success of our action — the actual achievement of our ends — is contingent on it being true that they will act as trusted. If a trustor merely has non-epistemic grounds to trust, they may, at most, be inclined to create conditions conducive to trust (such as acquainting themselves with the potential trustee or inquiring about their conduct). However, the appropriateness or justification of trust requires having epistemic grounds for trusting. It ultimately depends on one's assessment whether the trustee is in fact committed to acting as trusted, and capable of fulfilling their commitments. In other words, justifiable and rational trust require having reasons that pertain to the trustee to believe they are trustworthy. Without some assessment of trustworthiness, Longworth endorses a far too naive picture of interpersonal relationships, where agents consistently place trust in others, only to face repeated disappointment.

Setting aside the plausibility of Longworth's view, he acknowledges that his view does not encompass moral trust, stating it as a separate question that remains open:

> I wish to table [questions that] concern trust in other people's good disposition of the will, rather than in other competences of theirs on which we might rely. [...] Can we have sufficient evidence about the dispositions of others' will? (2017, 269)

Moral trust is a special case indeed, as this passage indicates, and it requires the assessment of the moral dispositions of others in ways that could provide grounds for their moral trustworthiness. However, given that for Kant's "Opacity Thesis" — the claim that we can't truly know one's grounds for moral action, and so we can't know their moral character and worth (see *GMS* 4:407; *R* 6: 51/71; cf. *MS* 6:447) — it is hard to think of sufficient grounds for moral trust.[4]

[4] Kant makes this claim both in the Groundwork and in the second Critique: "no certain example can be cited of the disposition to act from pure duty" (*G* 4:406), "It is absolutely impossible by means of experience to make out with complete certainty a single case in which the maxim of an action… rested simple on moral grounds (*G*

In addition to the problem of grounding moral trust, which is epistemic in nature, we must address first a prerequisite query preceding the examination of moral trust in others. Namely, before we take on the challenge of finding epistemic grounds for trusting others, we must ascertain the normative directives regarding the quality of being trustworthy, both in ourselves and our co-agents. That is, what is there in moral law that guides us to act in ways that are reliable and recognizable as conducive to engaging in trustful relations?

Thus, following Longworth's way of posing the question, we can elaborate on our Kantian account of moral trust, taking it to be grounded in other people's good dispositions of the will. Specifically, if trust consists of having confidence that an agent will ϕ, where ϕ is the action commanded by the moral law in a given circumstance, then moral trust is not merely believing that the agent will ϕ. Instead, it is one's entails a confidence that they will ϕ following their moral disposition — a confidence we get from our assessment of their moral conduct.

Let's illustrate how moral trust would play out using a couple of classic examples within Kant's scholarship. Take the prudent shopkeeper—an individual who behaves honestly not because that's the right thing to do but rather for prudential reasons. For instance, they may refrain from overcharging their customers, but only because it would hurt their reputation and business. This example is usually brought in the context of discussions of moral worth.[5] However, it also underscores the unreliability of the shopkeeper's actions based on the quality of their will. Their moral character is "only very contingent and precarious" (*GMS* 4:390), rendering them unreliable for consistent moral actions, and thus, not morally trustworthy.

4:407), and "no example of exact observance of [the moral law] can be found in experience (*KpV* 5:47). For a thorough analysis of how the opacity thesis concerns the assessment of other agents, see Berg (2020, sec. 2).

[5] See, Simmons (1989), Sliwa (2016) and Way (2017).

Now, think about the 'murderer at the door' scenario, the famous challenge Benjamin Constant posed to Kant's categorical morality, as described here:

> [Kant] goes so far as to maintain that it would be a crime to lie to a murderer who asked us whether a friend of ours whom he is pursuing has taken refuge in our house…. It is a duty to tell the truth. (*VRML* 8:425)

In his response to Constant's challenge, Kant seems to bite the bullet and deems lying to a murderer seeking their victim's whereabouts as immoral. Despite our intuitive resistance to an unwavering commitment to truth, Kant's response targets the idea that, when acting morally and telling the truth the agent faces no moral reproach because:

> If you had lied and said that he is not at home, and he has actually gone out (though you are not aware of it), so that the murderer encounters him while going away and perpetrates his deed on him, then you can by right be prosecuted as the author of his death. (*VRML* 8:427)

In this example, Kant suggests that the door answerer must still obey the moral law and tell the truth. He justifies this claim by arguing that if they lied and the friend had acted under the premise that they had told the truth, they could reasonably be held accountable for the friend's death. Notably, Kant does not assume that the friend would rely on the door answerer lying, although Constant (and many of us, in fact) take lying to be the only reasonable action in such a case. Instead, Kant thinks it's reasonable that the friend would act on the premise that the door answerer told the truth and revealed his location, making it the right decision (to the extent that it would keep him safe) to escape.

I believe we can make better sense of Kant's response to Constant's challenge if we think of it in terms of moral trust. Specifically, as if the friend held the door answerer as a morally

trustworthy agent and acted under such premise in their deliberation of how to proceed — rely on their lie and calmly wait in the living room, or trust their moral disposition and escape. I argue that trusting someone to act morally implies confidence that they will do so in any circumstance. In this scenario, the friend who has taken refuge in the house can be assured that his morally good friend, who had just opened the door and is talking to the murderer, will obey the moral law and tell the truth. He doesn't have to rely on contingent considerations that might guide them to act otherwise. Therefore, moral trust consists of having epistemic confidence in the trustee's adherence to the moral law, allowing the potential victim to act based on that confidence.

While these examples shed light on moral trust in the context of telling the truth, this perspective can be broadened to encompass moral conduct more broadly, especially to the extent that the trustee is morally committed to the trustor. Furthermore, this idea of moral trust treats trust, not mere reliance, by maintaining the vulnerability to their agential stance.

3. The Normative Grounds for Moral Trustworthiness

> *So act 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 a means*. (*G* 4:429)

The proposition that moral trust is directed at the trustee's morally right conduct towards us prompts an examination of *FH*, where he strives to bring the categorical imperative "closer to intuition" by applying the principle of morality to interpersonal interactions, as an attitude of respect for persons.[6] Given this emphasis on the morally constitutive relations among human

[6] Kant himself doesn't explicitly say that the Formula of Humanity is coextensive with the idea of respect for persons, but it can be confirmed in different ways, It doesn't only derive from passages where Kant contends that humanity is an object of respect and from the examples Kant provides for treating others as ends (e.g., refraining from making false promises). It is also prevalent in the literature, as prominent scholars like Allen Wood and Andrews Reath argue that claims of respect for persons are essential for the practical application of the Formula of Humanity in our deliberations(Reath 2013, 203; Wood 1999, 117), and others, such as O'Neill (1985) and Korsgaard (1992), advocate for prioritizing the co-agent's autonomous stance by demanding their

beings, I suggest we explore *FH* to unveil its potential implications for a Kantian comprehension of moral trust. First, I want to describe what normatively grounds the duty to treat others as ends and in what sense it constitutes a form of respect. Then, I examine its implications for trust to argue that what Kant identifies as the conduct of those who consistently treat others as ends constitutes trustworthiness.

Kant defines humanity in its practical sense as "the capacity to set oneself an end – any end whatsoever" (*MS* 6:392; *RGV* 6:27-28), emphasizing that rational beings are characterized by their free ability to choose the ends they seek to achieve through their actions. Humanity in a person manifests as the potential for this capacity to be directed towards the moral law, which Kant calls *personality*. In this perspective, human beings, as persons, possess a unique rational capacity.

Considering Kant's view of human nature as grounded in both reason and sensibility, *humanity* is rooted in reason rather than sensibility, which is susceptible to pathological desires. The 'use' of a person's humanity equates to the use of their agential capacities. "Using"' may initially appear problematic to those not familiar with Kant. However, in Kantian terms, "using" a person merely means to act on and with them, in circumstances where our principles align or depend on the actions of others. Hence, *FH* is an object of great interest, addressing the interpersonal dimensions of the categorical imperative. It not only sets criteria for morally permissible actions but directly establishes a standard for conduct toward ourselves and fellow human beings (Geiger 2023, 565).

Kant identifies the humanity in us as "the object of respect," in virtue of an inner worth that pertains to all rational beings as such:

fully informed and capacitated consent in our interactions with them, which is naturally a way of respecting our co-agents and giving primacy to their will over a joint action.

> But man regarded as a person […] is not to be valued merely as a means to the end of others or even to his own ends, but as an end in himself, that is, he possesses a dignity (an absolute inner worth) by which he exacts *respect* for himself from all other rational beings in the world. […] Humanity in his person is the object of respect which he can demand from every other man, but which he must also not forfeit. (*MS* 6:135, italics in original)

Following this passage, *FH* encapsulates the obligation to always regard others as ends in themselves, alongside the injunction against instrumentalizing them. Given our focus on the positive duty of respect, let's look more closely at the former aspect.[7]

According to Kant, an end is "what serves the will as the objective ground of its self-determination" (*G* 4:427). Persons, with their end in themselves, have their will grounded in their dignity. This dignity is what assigns value to their will, endowing their autonomy with intrinsic worth that demands respect and preservation. It is through this inherent dignity, this absolute inner worth, that the duty to show respect to both oneself and others arises, as they share a comparable intrinsic value.

Respect, in its practical sense, is a moral attitude derived from moral deliberation over its subject (S. Darwall 2008, 251). Similar to how respecting the law involves a decision to act morally out of duty, showing respect for persons—forming the foundation of our moral interactions with them—originates from practical reasoning and the acknowledgment of their inherent value.[8]

[7] In fact, for some Kantians, understanding Kant's position on treating people as ends in themselves is far more crucial than understanding the negative argument for treating people as mere means (Wood 1999, 143).

[8] This sort of respect is made necessary by a person's moral and rational nature and stands different from what Stephen Darwall calls "appraisal respect," which is the kind of respect that we may or may not hold towards other people based on certain contingent features and merits that they have, like, for example, being very good at baseball or in analytical thought (S. L. Darwall 1977).

Considering that *FH* stems from a duty to respect others, we can view this rendition of the categorical imperative as a specification of how to act upon our fellow human beings. It serves as a principle or rule of conduct that stems from respect. With that in mind, let's delve into Kant's understanding of that rule of conduct so that we can draw the appropriate conclusions about what constitutes the conduct of a trustworthy agent.[9]

Following Kant's exposition of *FH* in the *Groundwork,* Kant goes on to different scenarios where the formula may come to hand. One that emphasizes the use of humanity in others in the most useful way for us, representing a case of perfect duty, is the example of the false promisor:

> As regards necessary duty to others or duty owed them, he who has it in mind to make a false promise to others sees at once that he wants to make use of another human being merely as a means, without the other at the same time containing in himself the end. For, he whom I want to use for my purposes by such a promise cannot possibly agree to my way of behaving towards him, and so himself contain the end of this action. (*G* 4:429-430)

The principle works here in a twofold manner. First, the most evident and emphasized facet of interacting with someone as an end is consent. For us to uphold a morally appropriate relationship with our co-agents, we need their consent in some capacity. This consent condition has been underscored in various ways. Some contend that treating our fellow agents as ends in themselves requires their potential consent (O'Neill 1985; Korsgaard 1992).[10]

[9] This is not to say that producing actions similar to those of a trustworthy agent necessarily implies their trustworthiness. However, if such actions result from a maxim derived from the categorical imperative, and if this moral reasoning emanates from the moral disposition of that agent, then we may indeed conclude that a law mandating agents to act in trustworthy ways is one that renders them trustworthy agents.

[10] Scholars have criticized the possible-consent account thoroughly, not only for driving counterintuitive results in its moral deliberation (Kerstein 2009), but also for standing in tension with parts of Kant's political philosophy where instances of use of force on others are morally permitted (Pallikkathayil 2010).

Conversely, others advocate for actual consent (Kleingeld 2020) while some acknowledge the morally problematic settings in which consent is often given and thus emphasize what we might term rational or justified consent (Parfit 2011). I have no intention of delving into this debate in this paper. However, when we examine the positive aspect of this condition—specifically, not viewing the absence of consent as merely treating another as a means, but rather as creating space for consent as a means of treating them as ends in themselves—one point of agreement among all three approaches is their understanding of consent as a normative power intrinsic to agents. It is an act of granting permission and effecting a normative change in the world that cannot be attributed to mere means or objects (Fahmy 2023, 42).

The second facet of this principle comes to light when Kant asserts that for someone to be treated as an end, the other person must "contain in himself the end" of the action they are engaged in (*G* 4:430). This implies that when engaged in action with another person, a shared end in our cooperation must be possible — both parties decide to promote the same end in cooperation. In the case of the false promisor, however, their choice to make a false promise cannot garner approval from their co-agent, who ends up deceived. By resorting to a false promise, one prevents their co-agent from choosing their end, and so severs their participation from the potential of sharing the end that guides their collaborative action (Korsgaard 1996, 139).

Back to our discussion of trust, moral trust encompasses confidence in another person's moral disposition, marked by reliance grounded in their moral conduct. This entails a vulnerability to the trustee's agency, in which the trustor becomes susceptible to whether the trustee will decide to act morally. Following our analysis of *FH*, I argue that trustworthiness is constituted by the trustee's capacity and willingness to act in accordance with the moral law in their relationships with others, thus making themselves worthy of their trust.

This practical stance goes beyond mere reliance, imparting trust with its distinctive relational dynamics. While from a strictly moral standpoint, a break of trust falls into line with any other failure of respect, justifying the trustor in holding the trustee accountable for their behavior, trust occurs within the context of a trusting relationship. Thus, a break of trust feels distinctively personal and warrants feelings of betrayal or personal disappointment, going beyond mere moral disapproval.

In *The Metaphysics of Morals*, Kant says that persons "exact respect for [themselves] from all other rational beings" in virtue of their dignity, and that "humanity in [one's] own person is the object of the respect which [one] can demand from every other human being" (*MS* 6:434-435). Therefore, respect is something we may rightfully expect and demand from fellow human beings as moral agents. The mere fact that they are such establishes a *pro tanto* basis for relying on their moral stand towards us. While this doesn't suffice to establish moral trust *per se*, it constitutes a preliminary condition for engaging in moral trustful relations. This premise implies that we can demand others to act morally, and thus, in a trustworthy manner, with us.

As discussed earlier, *FH* serves as the foundation for human interactions grounded in respect, thus casting respect as a duty owed to all persons. A duty of respect is an ethical obligation directed toward specific ends or maxims regarding others. Respecting others, as Darwall aptly points out is "the maxim of limiting our self-esteem by the dignity of humanity in another person" (6:449) (S. Darwall 2008, 265).

This suggests that one owes it to the moral trustor to act out of respect toward them based on their humanity. It is a personal duty, rooted in the trustor's dignity. This duty forms the responsive practical stance we strive for in moral trust. It's a commitment undertaken in recognition of the inherent worth of the other person, aligning with the duty of respect as

defined by Kant. Furthermore, this responsive practical stance, when combined with others' demand for respect, fosters a mutual moral commitment, or as Darwall puts it, “it implies that it includes an authority to make claims and demands of one another, and so hold one another responsible as equals. […] It entails the further proposition that persons are mutually accountable for this treatment, that they must be able to justify their treatment of each other to one another” (S. Darwall 2008, 264).

Thus, each of the co-agents holds a valid claim to respect, shaping the foundation of their morally trustful interaction. Respect, akin to trust, is an asymmetrical relationship that may also manifest itself symmetrically in human interactions, wherein the parties take both an active role (acting upon) and a passive role (being acted upon). The trustor demands respect by virtue of their dignity, expecting the trustee to uphold the trustor's inherent worth. Simultaneously, the trustee must limit their actions according to the dignity of the other person, as prescribed by the moral law. This dynamic creates both a right and a duty, respectively, to which they are committed in their capacity as moral beings.

The authority the trustor wields in demanding respect results in dynamics similar to those of trust. In such dynamics, the trustee bears responsibility and accountability for any morally wrong actions, justifying their accountability for behavior that might potentially leave the trustor feeling betrayed or let down.[11] Failing to act morally in these cases is not simply a failure of character, but constitutes a violation of duty to the trustor, as it entails a failure to recognize and act on them as persons deserving of respect.

[11]While the violation of the duty to the trustor may already constitute betrayal, my focus here on the appropriateness of feelings of betrayal is because this is the relevant aspect of the phenomenology of trust in which we can distinguish it from reliance most clearly. Therefore, it is important to point out how it is enabled in a Kantian moral trust.

4. The Epistemic Puzzle of Morally Trusting Others

From the perspective of trust, there is something incomplete about *FH*. While it can establish a foundation for trustworthiness by guiding us to act in ways that can be relied upon and recognize others as beings with whom we could potentially engage in trusting relations, it doesn't tell us they will indeed fulfill that potential. For this, we need some reason to believe that they're not just generally subordinate to moral law as moral beings, but that they're genuinely committed to acting morally, and thus, in a trustworthy manner.

This question echoes Longworth's challenge in his final remarks, where he asks what are the reasons (if any) for trusting the moral disposition of others, and he rightly emphasizes that we lack access to the moral character of others (and this is true even in relation to ourselves, see *GMS* 4:407). While a morally good agent is inherently trustworthy, placing confidence in the quality of others' moral dispositions appears problematic. This presents a significant puzzle for constructing a plausible and justified account of moral trust.

It's worth highlighting that Kant himself expresses skepticism about trusting others, as evident in discussions on friendship in his *Lectures on Ethics*:

> We must so conduct ourselves to a friend, that it does us no harm if he were to become our enemy; we must give him nothing to use against us. We are not, indeed, to suppose that he may become our enemy, for then there would be no trust between us. But if we give ourselves entirely to a friend, and entrust him with all the secrets which might detract from our happiness, and might well be divulged if he did become an enemy, then it is very unwise to tell him these things, since he could either give them away through inadvertence, or use them to our hurt if he became our foe. (*VE* 27: 429–30)

Now, to me, the relationship described in this passage doesn't quite resemble trust, let alone friendship. Kant defines friendship as "the union of two persons through equal mutual love and respect" (*MS* 6:469), and thus, if my argument about what respect for persons entails is correct, Kant should allow for a more vulnerable form of trust — a moral trust, in fact — one that is based on the friend's recognition of our dignity, providing confidence that they will act morally even if the friendship comes to an end. Namely, if the friend respects us as persons and not merely as objects of their love, then we could trust them, disregarding the special relationship we hold with them.[12]

Thus, a Kantian conception of moral trust, of the kind we could potentially find in friendship, must accommodate and embrace vulnerability as an attitude of respect for others. Nevertheless, for the same reasons Kant is hesitant to make oneself vulnerable, we must look for reasons to believe the other person, a potential trustee, indeed respects us as persons, and thus, be justified in morally trusting them. Disregarding any partial relationships we may or may not hold with them, we must look for a way of assessing their moral conduct so that we can have confidence that they will act morally with us.[13]

[12] This has interesting implications for the very idea of friendship in Kant: For the morally virtuous agent, friendship is merely a relation of love, as respect is inherently present in all their interactions with fellow human beings. Similarly, the conclusion of a friendship, from their perspective, would only mean the discontinuation of love (at least from one of the parties involved), but not a cessation of respect.

[13]Moral trust, however, requires some basis for relying on their moral conduct. Therefore, an assessment of their moral conduct is not just one way of establishing moral trust but the only way. While this means that our trust is limited to their commitment to doing the morally right thing, it doesn't imply a lack of concern for our interests across the board. There is a basis to assert that the moral trustee can also be trusted to promote our ends more generally. Kant acknowledges duties of beneficence that command the moral agent not only to treat others with respect by not using them as mere means but also to act in ways that develop and promote others' ends, particularly fostering their autonomous capacities (*GMS* 4:430). Nevertheless, our levels of reliance on the trustee in relation to these duties will inevitably be lower, as they constitute imperfect duties, and it is at the discretion of the trustee whether to act on them. For a more in-depth exploration of duties of beneficence, see Cummiskey (1996).

5. An ‘Ideal’ Solution to the Puzzle

I’d like to propose here a solution to the puzzle, or a basis for moral trust in Kant. In exploring how we might be able to assess one’s moral conduct in spite of the opacity of their motives, I turn to Korsgaard's notion of “a neighborhood of the kingdom of ends,” which helps us envision an ideal relationship between two moral agents. This approach aims to facilitate a comparative assessment of another person’s moral conduct in relation to this ideal standard.

Korsgaard, in her book *Creating the Kingdom of Ends,* lays the grounds for morally trusting other people saying that:

> To hold someone responsible is to regard her as a person — that is to say, as a free and equal person, capable of acting both rationally and morally. It is therefore to regard her as someone with whom you can enter the kind of relation that is possible only among free and equal rational beings: a relation of reciprocity. When you hold someone responsible, you are prepared to exchange lawless individual activity for reciprocity in some or all of its forms [...] you are ready to trust, and generally speaking to risk your happiness or success on the hope that she will turn out to be human (1992, 188–90)

For Korsgaard, holding someone morally responsible involves viewing them as a person—a morally apt being capable of moral deliberation and action. Considering others as moral agents sets the stage for various forms of reciprocal relationships, with trust being a central one. She borrows this idea of reciprocity as foundational for interpersonal relations from Kant's definition of friendship, as a perfect reciprocal relation, rooted in the fundamental attitudes of our human nature—love, an attitude of inclination towards the friend, stemming from our sensible nature; and respect for them as persons, stemming from our rational nature.

In this context, Korsgaard argues that personal relations, where individuals acknowledge each other as morally responsible beings, allow for reciprocal relationships grounded in mutual respect.

This recognition has the potential to create what she calls "a neighborhood where the kingdom of ends is real" (Korsgaard 1992, 194).[14] Similar to the idea of the kingdom of ends as a "systematic union of various rational beings through common laws" (*G* 4:434), where we envision societal cooperation at its best with citizens acting morally with one another under universal laws, a 'neighborhood of the kingdom of ends' envisions a local, interpersonal cooperation, where two individuals act morally with one another under universal laws. In such settings, we may also include the idea of moral trust, as it derives from the moral conduct of the agents involved.[15]

Drawing on Korsgaard's idea, I propose a local version of the kingdom of ends that focuses on how agents act upon each other and establish relations of moral trust. Although I borrow this idea from Korsgaard, I diverge from her with respect to the roles of reciprocity and responsibility in these interpersonal settings.

First, moral trust doesn't derive from reciprocity in any way. While trusting relationships involve reciprocity, moral trust, as discussed above, is an asymmetrical attitude because the morally good trustee will act morally independently of my expectations or actions toward them. One doesn't make themselves trustworthy by virtue of a reciprocal relationship, but

[14] This quote from Korsgaard describes a relation of friendship. Nonetheless, I take the liberty to use it here to describe merely the idealized moral communion between persons, as Korsgaard herself says that friendship encompasses more than what is required for a kingdom of ends. While friendship requires a combination of mutual love and respect in their maximal form, the kingdom of ends requires only respect.

[15] I intentionally avoid here the discourse of social and political relations commonly linked with the concept of the kingdom of ends. In taking this stance, I align myself with Korsgaard's perspective, which considers interpersonal relations involving two or a few more people as part of a broader continuum that extends to more institutionalized networks.

solely by virtue of their own commitment to acting morally. Similarly, one may have moral trust in another by virtue of their assessment of the trustee's moral conduct while not being themselves committed to acting morally (as when one morally trusts for mere prudential reasons). This of course doesn't dismiss the trustor's moral obligation to act in a trustworthy manner, but it highlights the autonomous nature of the trustee, and underscores that reciprocity can't be a prerequisite for acting trustfully. If it were, one's trustworthy conduct would be responsive to another's conduct, eroding its foundation in respect for persons.

Second, merely holding someone responsible is insufficient for establishing moral trust, contrary to what Korsgaard seems to imply. While she emphasizes the latter as the condition for entering relations of reciprocity such as trust, Kant, as discussed earlier, doesn't advocate for a naive stance in our engagements with others, as it is evident in his cautious approach to trusting even friends. Responsibility, assuming a Kantian framework, stems from taking the other as a moral agent, one capable of acting morally and from whom we can demand respect, but it doesn't mean we can justifiably "regard [them] as someone with whom you can enter the kind of relation that is possible only among free and equal rational beings" (Korsgaard 1992, 189). For engaging in trusting relations, we need more than moral aptitude that entails responsibility — we need to take them to be actively committed and willing to act morally with us.

Furthermore, acknowledging the other person as an autonomous moral agent, who may or may not choose to act morally, is integral to recognizing them as ends in themselves. This acknowledgment underscores trust's vulnerability, as it is contingent upon the will of the other person. In this regard, Longworth's exploration of interpersonal trust within Kant's account of trust holds merit. While I claim that mere practical non-epistemic merits are insufficient to establish moral trust, Longworth is right to say that to accommodate trust we must let go of certainty. Akin to Kant's famous claim that we must "deny knowledge to make

room for faith' (B XXX), for the sake of advancing our ends in an interpersonal and social environment, we must make room for assessing others' conduct even though we cannot really know whether it reflects their moral disposition.

Thus, we need a conceptual framework for understanding moral trust that allows us to go beyond the minimal requirement of responsibility, assessing moral conduct for trusting our co-agents. A local "neighborhood of the kingdom of ends" seems to fit here as a representation of interpersonal cooperation, where two individuals act morally with one another under universal laws. This idea we borrow from Korsgaard becomes especially useful if we take it to be an ideal, a standard of conduct, as Barbara Herman suggests.

Herman, by tapping into Rawls' focus on the kingdom of ends as an ideal with a role in judgment (*GMS* 4:433), argues that the latter acts as a standard against which we can make judgments about social and interpersonal contexts. I argue that this form of comparative assessment, embedded in the explicit sociality of the Kingdom of Ends, is promising in the task of assessing the conduct of our co-agents, and thus, whether we can rationally trust them to act morally.

Herman draws upon Kant's remarks on transcendental ideals from the first *Critique* to elucidate the notion of the kingdom of ends as a guiding model for judgment:

> The general concept of the ideal is introduced this way. 'No objects can be represented through pure concepts of the understanding apart from the conditions of sensibility.' When applied to appearances, the pure concepts of the understanding — the categories — "can be exhibited *in concreto*," This is so because ideas "contain a certain completeness to which no possible empirical knowledge ever attains.' […] Reason, rather, 'thinks for itself an object which it regards as being completely determinable in accordance with principles' Although the ideal cannot exist, even in

> example, it is not 'a figment of the brain.' The ideal supplies reasons with a standard of judgment. (Herman 1997, 200)

According to this framework, ideals serve as representations determined by practical principles of reason, guiding our conduct toward a form of perfection. They wield practical influence by acting as a framework that underscores the proximity of actions or attitudes to perfection. Herman places particular emphasis on the use of such ideals as a way around Kant's Opacity Thesis, proposing we use the kingdom of ends for assessing the quality of one's moral conduct, and so their moral worth, by comparing their vices and faults to the ideal of moral perfection. In this way, ideals supposedly facilitate moral self-assessment and improvement. This suggestion is further supported by the following passage from Kant's *Lectures on Ethics*:

> [practical ideals] are not chimeras, for they constitute the guideline to which we must constantly approach. [...] a yardstick by which to estimate our moral worth, and to know the degree to which we are faulty and deficient; and here I have to conceive of a maximum, so that I know how far away I am, or how near I come to it. (*VE* 29:604-605)[16]

Although this perspective primarily applies the kingdom of ends to self-assessments, it is equally adept at evaluating the moral conduct of others, functioning as a standard for moral conduct by which we can 'compare and judge' (A569/B597) the moral conduct of our fellow agents, given that they are also represented in this ideal.[17] The kingdom of ends represents a

[16] Englert (2022) brings in this passage in a recent paper, to make a point similar to that of Herman, characterizing practical ideals as instrumental in evaluating moral deficiencies.

[17] To provide some necessary context, in the passage from the first Critique that I'm referencing, Kant argues that practical ideals may serve us to "compare and judge ourselves, and so reform ourselves." The latter component, of self-moral reform, isn't relevant to the assessment of others, as we can't engage in reforming their moral character.

social order where agents, bound by the moral law, interact, providing a standard for comparative assessment of moral conduct against the same perfect standard.

When determining whether another person can justifiably be trusted, we assess their conduct in comparison to a standard of moral trustworthiness. I argue this form of comparative assessment provides us with appropriate (subjective) grounds to trust them, and it ultimately reflects our confidence in their moral conduct. The trustee's moral disposition is opaque to us, much like our own, but looking at their conduct in comparison to its ideal form, we get a "yardstick by which we estimate [*their*] moral worth, and to know the degree to which [*they, i.e., the trustee*] are faulty and deficient."

Within such an ideal of interpersonal relationships, marked by perfect adherence of all members of the kingdom of ends to the moral law, those can be rightfully (and still, only in our thoughts) regarded as trustworthy. Within the kingdom of ends, the ideal trustee serves as an exemplar through which we may assess the trustworthiness of fellow agents in the real world. By comparing their conduct to the exemplar of trustworthiness, we discern the proximity of their conduct to that of a member of the kingdom of ends, helping us decide whether we can place moral trust in them. For instance, if we return to the friend example from the beginning, we can morally trust our friend by comparing their conduct to that of the ideal agent—one who is capable of calling you out in the face of a moral lapse, keeping promises even to those with whom they no longer have personal ties, and so on. The local kingdom of ends, encompassing the ideal moral counterpart, serves as the ruler, and we can measure the friend, considering their past actions and expressions of character traits, in light of it.

Now, there is an apparent tension between the possibility of assessing others' moral conduct as it becomes necessary for trust and Kant's claim that we cannot know one's moral

disposition. In other words, how can we compare what is opaque to us? This tension is resolved by distinguishing the kind of assessment necessary for justified moral trust. The comparison we are doing is between actions, and not dispositions. Namely, we compare the overall conduct of potential trustees with the conduct of ideal ones. By definition, we only know that the latter is grounded in a perfect moral disposition, and we cannot know the same about the former due to the opaqueness of their motives. Nevertheless, I argue it is plausible to take one's overall conduct — when compared with that of an ideal moral trustee — to serve as some indication that it is also grounded in a good (albeit not perfect) moral disposition. Of course, the potential trustee can be only acting according to duty and not out of respect of duty, but if we make an overall assessment — based on a sufficiently general poll of their actions and external expressions of character traits — and find reasonable compatibility with the conduct of the ideal trustee, that must count for something. Although such comparison would never be conclusive, and we can never claim knowledge or certainty about the trustee's motives, we still find some epistemic grounds for morally trusting them. In this way, the ideal of a moral trustee provides epistemic merit for moral trust. It allows us to assess whether it is more likely to be true that the trustee will act out of a good moral disposition, rather than basing our trust merely on non-epistemic practical merits.

In an act of moral trust, we embrace the vulnerability that comes with acknowledging our fellow agents as truly free, capable of disappointing us in moral failure despite their proximity to the ideal represented in the kingdom of ends. Moral trust is warranted when we believe our fellow agents will act morally. This form of trust maintains the delicate balance between acknowledging the potential for breaches of trust, while allowing us to have rational and justified confidence that our fellow agents will honor their moral duties.

This method offers grounds for moral trust that aligns with how Andrew Chignell (2007) describes a mere Conviction in Kant's epistemology — i.e., conviction that doesn't amount to

knowledge. According to Chignell's reading of Kant, knowledge is a true assent that is objectively and subjectively grounded. That is, it is grounded on "experiences and/or assents that… licenses [us to] assent with a moderate-to-high degree of confidence" and in addition, the subject is in a position to cite those objective grounds upon reflection, respectively (Chignell 2007, 327–28). A mere conviction, on the other hand, is an assent that has sufficient objective grounds, but where the subject is "not in a position, even on reflection, to cite those grounds" (2007, 332). Chignell takes objective grounds to refer to "perceptual, memorial, and introspective states, as well as other sufficient assents we already hold (the results of inductive and deductive arguments, assents about what others have testified, assents about one's experiences, and so forth" (Chignell 2007, 327).

When rationally establishing moral trust through the method I lay out here, we get a similar picture: the trustor refers to their observations of the trustee's conduct and compares them to an ideal that is indeed grounded in a good moral character. If one grounds trust sufficiently well, referring to enough evidence of good moral conduct (in actions and expressions of character traits that align with the ideal trustee's conduct), it is plausible to say that they may trust another with a moderate-to-high degree of confidence. Nevertheless, due to the opacity of motives that Kant emphasized, the trustor is never in a position to directly access or cite the moral disposition that grounds another's trustworthy actions, leaving them short of subjective grounds.

In this way, we are able to reach rational confidence while acknowledging our limits in accessing the trustee's moral disposition. Chignell rightly notes that Kant is a fallibilist about sufficient objective grounds, holding that one can have a rationally-established conviction, for example, of the form "I can trust her," and the assent still may turn out to be false (2007, 330). This fallibility aligns with the vulnerability inherent in trust my sole point of agreement with Longworth — that we must give up on knowledge to make room for trust. Furthermore,

it contains a fundamental difference between our accounts. For Longworth, trust consists is akin to a practical belief (*Glaube*), and therefore finds sufficient grounds in mere non-epistemic merits found in social cooperation. In my account of moral trust, on the other hand, I suggest we conceptualize trust based on a conviction about the trustee's moral character. This requires finding sufficient epistemic grounds that they will act morally. Although we cannot fully cite those grounds, they still seem to give us a concrete indication of the trustee's conduct, avoiding a naive or overly optimistic account.

6. Conclusion

In this paper, I present a Kantian perspective on moral trust, built upon two fundamental pillars. First, I argue that moral virtuosity implies moral trustworthiness. This assertion arises from the recognition that all human beings possess a fundamental dignity, instilling a duty to act with respect toward them. Acting morally, in this context, equates to acting in a trustworthy manner – conduct that can be relied upon and is rooted in a commitment to respecting the trustor as a person. Second, I argue that the concept of the kingdom of ends, when envisioned locally, focusing on an interpersonal relationship, offers a way to assess the moral conduct of our fellow agents in comparison to their ideal moral form. This allows us to establish trust in others in a nuanced way, having possible grounds for holding them trustworthy while avoiding any aspirations for certainty that would undermine their agency.

One of the strengths of this account is that it situates the role of trust in Kant's moral project. Contrary to what Korsgaard argues, trust isn't dependent on reciprocity, but is a duty that promotes it. Moral trust can, of course, be used in bad faith when someone morally trusts another's moral behavior but doesn't treat them as ends in themselves. In such cases, the trustor transgresses the categorical imperative and also violates the natural expectations of reciprocity that motivate us in our relationships (and Korsgaard in her account). This,

however, doesn't pose a challenge to the very idea of moral trust: The trustworthy agent acts morally, disregarding any malicious use of their actions. One's moral trustworthiness can be taken advantage of. Still, contrary to other forms of trust, the moral value of the actions performed by the trustee isn't jeopardized by such a lack of reciprocity.

Finally, by providing a moral approach to engaging in trusting relationships, my account not only introduces the possibility of moral trust but also asserts the duty of being morally trustworthy — akin to the duty of being respectful of other human beings. This duty is evident in the passage from *The Metaphysics of Morals* cited in the introduction of this paper, where Kant claims we have ought to utilize our moral perfections in social interactions and to "cultivate a disposition of reciprocity—agreeableness, tolerance, mutual love, and respect" (*MS* 6:473). In the account laid out here, I see a true form of 'using one's moral perfection in social intercourse,' where we ground trusting relationships not in contingent commitments to one another but on the one categorical commitment common to all.

References

Kant's Works

Citations appear in the order of abbreviation, and page number from the Akademie Ausgabe, Kants Gesammelte Schriften, edited by Königlich Preussische Akademie der Wissenschaften (29 vols. Berlin: de Gruyter, 1900–). The only exceptions to this rule are in quotations from the Critique of Pure Reason, in which I follow the standard practice of referring to the first edition of 1781 (A) and the second edition of 1787 (B). All translations come from The Cambridge Edition of the Works of Immanuel Kant, edited by Paul Guyer and Allen Wood (Cambridge: Cambridge University Press, 1992–), and are located by the following abbreviations:

GMS, Groundwork of the Metaphysics of Morals. Translated by Mary Gregor;

KpV, Critique of Practical Reason. Translated by Mary Gregor;

MS, The Metaphysics of Morals. Translated by Mary Gregor;

RGV, Religion Within the Boundaries of Mere Reason. Translated by George Di Giovanni;

VE, Lectures on Ethics, translated by Peter Heath and J.B. Schneewind;

VRML, On a supposed right to lie from philanthropy. Translated by Mary Gregor.

Other Works

Baier, Annette. 1986. "Trust and Antitrust." *Ethics* 96 (2): 231–60. https://doi.org/10.1086/292745.

Berg, Anastasia N. A. 2020. "Kant on Moral Self-Opacity." *European Journal of Philosophy* 28 (3): 567–85. https://doi.org/10.1111/ejop.12520.

Chignell, Andrew. 2007. "Belief in Kant." *The Philosophical Review* 116 (3): 323–60.

Cummiskey, David. 1996. "The (Not So) Imperfect Duty of Beneficence." In *Kantian Consequentialism*, edited by David Cummiskey, 0. Oxford University Press. https://doi.org/10.1093/0195094530.003.0006.

Darwall, Stephen. 2008. "Kant on Respect, Dignity, and the Duty of Respect." In *Kant on Respect, Dignity, and the Duty of Respect*, 175–200. De Gruyter. https://doi.org/10.1515/9783110209655.175.

Darwall, Stephen L. 1977. "Two Kinds of Respect." *Ethics* 88 (1): 36–49.

Englert, Alexander T. 2022. "How a Kantian Ideal Can Be Practical." *Inquiry* 0 (0): 1–28. https://doi.org/10.1080/0020174X.2022.2075449.

Fahmy, Melissa Seymour. 2023. "Never Merely as a Means: Rethinking the Role and Relevance of Consent." *Kantian Review* 28 (1): 41–62. https://doi.org/10.1017/S136941542200053X.

Geiger, Ido. 2023. “Humanity and Personality – What, for Kant, Is the Source of Moral Normativity?” *Inquiry* 66 (4): 565–88. https://doi.org/10.1080/0020174X.2020.1779124.

Herman, Barbara. 1997. “A Cosmopolitan Kingdom of Ends.” 1997. https://philpapers.org/rec/HERACK.

Holton, Richard. 1994. “Deciding to Trust, Coming to Believe.” *Australasian Journal of Philosophy* 72 (1): 63–76. https://doi.org/10.1080/00048409412345881.

Jones, Karen. 1996. “Trust as an Affective Attitude.” *Ethics* 107 (1): 4–25. https://doi.org/10.1086/233694.

Kerstein, Samuel. 2009. “Treating Others Merely as Means.” *Utilitas* 21 (2): 163–80. https://doi.org/10.1017/s0953820809003458.

Kleingeld, Pauline. 2020. “How to Use Someone ‘Merely as a Means.’” *Kantian Review* 25 (3): 389–414. https://doi.org/10.1017/S1369415420000229.

Korsgaard, Christine M. 1992. “Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations.” *Philosophical Perspectives* 6:305. https://doi.org/10.2307/2214250.

Longworth, Guy. 2017. “Faith in Kant.” In *The Philosophy of Trust*, edited by Paul Faulkner and Thomas Simpson, 0. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198732549.003.0015.

Marušić, Berislav. 2017. “Trust, Reliance and the Participant Stance.” *Philosopher’s Imprint* 17 (17). http://hdl.handle.net/2027/spo.3521354.0017.017.

McLeod, Carolyn. 2021. “Trust.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2021. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entriesrust/.

O’Neill, Onora. 1985. “Between Consenting Adults.” *Philosophy & Public Affairs* 14 (3):

252–77.

———. 2002. *Autonomy and Trust in Bioethics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511606250.

Pallikkathayil, Japa. 2010. “Deriving Morality From Politics: Rethinking the Formula of Humanity.” *Ethics* 121 (1): 116–47.

Parfit, Derek. 2011. *On What Matters*. 1st edition. 2 vols. Oxford: Oxford University Press.

Reath, Andrews. 2013. “Formal Approaches to Kant’s Formula of Humanity.” In *Kant on Practical Justification: Interpretive Essays*, edited by Mark Timmons and Sorin Baiasu, 0. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195395686.003.0010.

Schröder, Peter. 2010. “26. ‘Irgend Ein Vertrauen … Muss … Übrig Bleiben’: The Idea of Trust in Kant’s Moral and Political Philosophy.” In *Cultivating Personhood: Kant and Asian Philosophy*, 391–400. De Gruyter. https://doi.org/10.1515/9783110226249.3.391.

Simmons, Keith. 1989. “Kant on Moral Worth.” *History of Philosophy Quarterly* 6 (1): 85–100.

Sliwa, Paulina. 2016. “Moral Worth and Moral Knowledge.” *Philosophy and Phenomenological Research* 93 (2): 393–418. https://doi.org/10.1111/phpr.12195.

Sussman, David G. 2001. *The Idea of Humanity: Anthropology and Anthroponomy in Kant’s Ethics*. Taylor & Francis.

Way, Jonathan. 2017. “Creditworthiness and Matching Principles.” In *Oxford Studies in Normative Ethics, Vol 7*, edited by Mark Timmons. Oxford University Press.

Wood, Allen W. 1999. *Kant’s Ethical Thought*. First Edition. New York: Cambridge University Press.