



This is a repository copy of *The difficult choices of trustworthy people*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/220864/>

Version: Published Version

---

**Article:**

Shemmer, Y. [orcid.org/0000-0001-5842-8470](https://orcid.org/0000-0001-5842-8470) (2025) The difficult choices of trustworthy people. *Philosophy & Public Affairs*, 53 (1). pp. 4-36. ISSN 0048-3915

<https://doi.org/10.1111/papa.12277>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:


<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

YONATAN SHEMMER\* 

# The Difficult Choices of Trustworthy People

## I. PREFACE

You are asked by your good friends to babysit their baby daughter for 2 h while they finish some urgent errands. “No problem” you say, “I am free for the next two hours, but don’t delay, I must catch a train in two and a half hours to reach an important meeting.” The allotted time elapses and there is no sign of your friends coming back. You find yourself in discretionary circumstances: the situation is not as you anticipated it would be when you agreed to babysit, and what you are meant to do in this situation was not specified when they entrusted her to you. What is the trustworthy thing to do? Miss your important meeting? Ask the new neighbor, whom you know, but not that well, to keep your friends’ baby until they come back? Take her with you to London?

## II. INTRODUCTION

This paper is about trustworthiness.<sup>1</sup> More specifically, it is about one aspect of trustworthiness, namely its discretionary nature. It concerns the

I am grateful to audiences at “The Normativity and Epistemology of Friendship” conference at Trinity College, Dublin and at the Annual Conference of the Israeli Philosophical Association for their comments and suggestions. Many thanks to Tsachi Keren-Paz for his guidance and advice regarding the legal literature on trusts. Many thanks also to an associate editor for *Philosophy & Public Affairs* whose advice and suggestions have greatly improved this paper. A special thank you to the following people for their extremely helpful suggestions and criticism: Chris Bennett, Graham Bex-Priestley, Novenka Bex-Priestley, Sarah Durling, Paul Faulkner, Daniel Schwartz, Ariana Shemmer, Ilya Shemmer, and Ruth Weintraub. If it were not for your trustworthy advice this paper would have been very different than the one I wanted it to be.

1. Many discussions of trust focus on the trusting side of the trust relation. They ask what it is to trust, what is the trusting attitude, or when is trust justified. Some aspects of these discussions of trust entail a view of what trustworthiness is. Other aspects merely suggest, as a natural accompaniment, certain views of trustworthiness. In the following footnotes I mention both authors that explicitly discuss trustworthiness and authors whose view of trustworthiness is merely suggested by their view of trusting; I designate the latter with an “\*.”

© 2024 The Author(s). *Philosophy & Public Affairs* published by Wiley Periodicals LLC. *Philosophy & Public Affairs* 9999, no. 9999

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

question of how to keep trust in discretionary circumstances. I will, for the most part, have little to say about other central questions answered by accounts of trustworthiness.<sup>2</sup> Indeed, the view I defend here is, as far as I can see, compatible with most of the available answers to these questions. Nevertheless, I think that all existing theories of trustworthiness, and by extension all theories of trust, suffer from an important lacuna. They underspecify what trustworthiness requires in discretionary circumstances.

In section two, I explain the discretionary nature of trustworthiness. In section three, I present five possible standards of trustworthy action in discretionary circumstances. I reject four of these standards. In section four, I consider an initial formulation of the fifth standard. Roughly, this

2. E.g.,

- Is reliability in intending to keep trust a necessary condition for trustworthiness? Katherine Hawley, *How to Be Trustworthy* (Oxford University Press, 2019), 73, 80; Karen Jones\*, “Trust as an Affective Attitude,” *Ethics* 107, no. 1 (1996): 9.
- Must the trustworthy action be motivated by a direct concern for the wellbeing or interest of the trustor, or could that concern be indirect? Jones\*, “Trust as an Affective,” 9.; Russel Hardin, *Trust & Trustworthiness* (Russell Sage Foundation, 2002), 24.
- Must the motivation that drives the trustee’s actions have its source in the realization that she is being trusted? Karen Jones, “Trustworthiness,” *Ethics* 123, no. 1 (2012): 68; Paul Faulkner and Thomas Simpson, eds., *The Philosophy of Trust* (Oxford University Press, 2017) 8; Annette Baier\*, “Trust and Antitrust,” *Ethics* 96, no. 2 (1986): 235.
- Must the trustee be motivated in any particular way or is it enough that she reliably intends to keep her obligations, or to keep her commitments? Baier\*, “Trust and Antitrust,” 254; Jones\*, “Trust as an Affective,” 4, 6; Carolyn McLeod\*, “Trust,” In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta and Uri Nodelman, Accessed Fall 2021 (Stanford University, 1997), <https://plato.stanford.edu/archives/fall2021/entries/trust/>; Christoph Kelp and Mona Simion, “What Is Trustworthiness?” *Nous* 57 (2023): 667–83. <https://doi:10.1111/nous.12448>; Katherine Hawley, “Trust, Distrust and Commitment,” *Nous* 48, no. 1 (2014): 16.

standard requires that the trustee be guided by the counterfactual wishes of the trustor for the discretionary circumstances. I discuss two families of objections—that the fifth standard is too onerous for the trustee, and that it is hostage to the possible irrationality of the trustor—and refine it in light of these objections. In section five, I take a step back to discuss two theoretical questions. First, what my account entails regarding the type of discretion that exists in relations of trust. Second, whether epistemic difficulties undermine the account. In the final section, I conclude and make a preliminary suggestion about the implication of my conclusion for accounts of trusting (as opposed to trustworthiness).

### III. SECTION TWO

My interest is in a feature of trust which has been emphasized by Annette Baier: that trust is a discretionary relation.<sup>3</sup> I will have more to say about what it means to be a discretionary relation, but for now let me highlight one aspect of relations of trust that make them discretionary, and which is compatible with all existing accounts of trust and trustworthiness. When a trustee is entrusted with a certain task, it is impossible for the trustor to fully specify how the task is to be completed. This is so because it is impossible to list all the ways in which the circumstances encountered by the trustee might differ from the circumstances anticipated by the trustor.<sup>4</sup> As a result, there is more to acting in a trustworthy way than following the

3. Baier, "Trust and Antitrust," 237. Annette Baier, "Trust," *Tanner Lectures on Human Values* 13 (1991), University of Utah Press, 117. Jones briefly mentions the discretionary nature of trust, but for the most part, this important insight of Baier has been ignored in the philosophical literature. Jones, "Trust as an Affective," 8.

4. As an anonymous referee has rightly pointed out, there are other reasons for the discretionary nature of trust. I have discussed above the fact that circumstances might differ from the ones anticipated by the trustor. To that we should add the fact that trustors, by necessity, anticipate circumstances in an underspecified way. It is impossible to imagine any circumstance in all its minute details, and it is impossible to describe all these minute details to the trustee. Furthermore, given the inescapable vagueness of linguistic expressions, the instructions given by the trustor, are necessarily vague. Such vagueness opens up the space for additional discretion. Finally, as I point out below, much entrusting occurs implicitly and thus the exact nature of what has been entrusted is left in yet another way for the trustee to determine. As the referee aptly put it, discretion "is the norm rather than the exception." In a different paper, I discuss the implication of these wider types of discretion for our understanding of trustworthiness. But for my current purpose the type of discretion discussed in the text will suffice.

instructions, explicit or implicit,<sup>5</sup> given by the trustor. There always exists the possibility that the trustee will have to make a decision about how to act in circumstances not covered by these instructions.<sup>67</sup> Consider the following example:

Limor (the Trustor) entrusts Deedee (the Trustee) with delivering a package to the Clarkes at 39 Marlborough Rd.

All might go as anticipated. Deedee might arrive to her destination and hand the package to one of the two adults of the Clarke family. But things might not go as planned. Marlborough Rd. might be blocked. The Clarkes might have moved. The Clarkes might have moved to 45 Marlborough Rd., they might have moved to the other side of town, or they might have moved to Dubai. The Clarkes might refuse to accept the package. The door might be opened by a house-sitter who promises to give them the package when they come back. Alternatively, the door might be opened by someone who refuses to say who she is and who looks like an intruder. A trembling child might open the door and tell Deedee that her parents were supposed to come back last night but did not. The list of possible unanticipated situations is endless. The list of further circumstances that might affect what to do in these various situations is also endless. It might be sunny on the day of delivery, but it might also be the worst storm of the year. The buses might work, but they might be on strike. Deedee might not have any plans for the rest of the day, but she might also be in a rush to meet her husband in the hospital.

Limor cannot think of all the possible situations that would prevent the entrusted act from being completed as planned, and certainly cannot think of all the possible wider circumstances that are relevant to a decision about what to do if it cannot be completed as planned. Saying that Limor cannot think of all these situations and circumstances is not a hyperbole. She literally cannot. No one can.

To be sure, Limor could have a single rule for all discretionary circumstances. As the next example shows, in most cases of entrusting, this

5. Much, if not most, entrusting is implicit. Even when the act of entrusting is explicit, exactly what is being entrusted is often determined implicitly, in part by contextual factors. There is, however, no harm in working with examples where the act of entrusting is, at least in part, explicit. See Baier, "Trust and Antitrust," 240.

6. One can enter relations of trust without an explicit or implicit *act* of entrusting (Hardin, *Trust & Trustworthiness*, 65; Baier, "Trust and Antitrust," 234).

7. In the remainder of the paper, whenever I discuss discretionary circumstances, it will be stipulated that it is impossible for the trustee to contact the trustor to ask for further instructions.

wouldn't be a good idea. Limor could tell Deedee that if for whatever reason the package cannot be delivered, she should e.g., bring it back to her. But consider the following scenario. Deedee puts the package in the back of her car and sets off for the Clarkes. On her way, she stops at the store. As she walks back to the car, she sees that it was broken into and that a bomb with a timer was attached to the package. Given the time left, she wouldn't be able to reach the Clarkes, but she does have just enough time to get it back to Limor, seconds before it would explode. . . To be clear, the point here is not that if Limor had specified that she wanted the package back in all discretionary circumstances, it would be untrustworthy to follow her instructions. The point is rather that any attempt to give such *catch-all* instructions would yield undesired results in some circumstances.

So far, I have noted that it is impossible to fully specify what one is to do if the task one was entrusted with cannot be completed as planned. Call this the *in-eliminability of (the possibility of) discretionary circumstances*.

Baier thinks that this in-eliminability teaches us something important about the attitude of trusting and about what it is to be a virtuous trustor. To trust, Baier thinks, is in part to realize that the trustee might face discretionary circumstances. And to be a virtuous trustor is, in part, to be forgiving, generous and open minded in one's reactive attitudes toward the decisions taken by the trustee in such circumstances.<sup>8</sup> This virtue helps the relation of trust flourish. The corresponding vice undermines it.

I believe that Baier's insight is crucial to the understanding of trust. I will not defend this claim here. As I said above, my primary interest is not in the nature of trusting but rather in the nature of trustworthiness.<sup>910</sup>

To see what the ineliminability of discretionary circumstances teaches us about trustworthiness, I propose we ask the following question:

What principles, or standards, characterize trustworthiness in discretionary circumstances? In other words, I propose we ask what a person should do in discretionary circumstances in order to be trustworthy.

8. Ibid. at 238.

9. As we will see at the end, my conclusions also have implications for the nature of trusting. However, to see these, we need to first inquire into the nature of trustworthiness.

10. Baier says that not anything goes in discretionary circumstances. She does not, however, give an account of the standards governing action in these circumstances. Ibid. at 237.

## IV. SECTION THREE

Suppose you find yourself in discretionary circumstances. The task you were entrusted with cannot be completed—at least not as it was explicitly or implicitly specified by the trustor. What should you do in these circumstances in order to be trustworthy?

I will consider five principles:

- (1) Any action is as trustworthy as any other action.
- (2) Try to complete the task entrusted to you as best you can.
- (3) Act as morality requires of you.
- (4) Act so as to maximize the well-being of the trustor.
- (5) Act as the trustor would have wanted you to act, in order to achieve his/her aim, had they anticipated the discretionary circumstances.

The existing philosophical literature is of little help in choosing among these five options. Even Baier, who has emphasized the discretionary nature of trust, says nothing about the principles which guide trustworthiness in discretionary circumstances.<sup>11</sup> Most other philosophers of trust don't even discuss the discretionary nature of this relation, let alone the principles that should govern action when discretionary circumstances arise.<sup>12</sup> Nevertheless, some views of trustworthiness may be interpreted as offering an implicit answer to our question.

On some views, all there is to trustworthiness is being disposed to fulfill your obligations<sup>13</sup> or your commitments.<sup>14</sup> You might think that such views implicitly offer the following answer to our question: in discretionary circumstances, make sure to fulfill your obligations or commitment with regard to these circumstances.<sup>15</sup> But this cannot be a

11. However, at times, it seems as though if she were to select a principle of trustworthiness in discretionary circumstances it would be principle #4. Baier\*, "Trust," 117-8.

12. Outside philosophy, discretionary decision making is considered at length in a variety of domains. One of them in particular, namely fiduciary law, might seem highly relevant to my current topic. I briefly discuss some of these domains toward the end of the current section and explain why the contribution of the associated literature to the analysis of trustworthiness is at most indirect.

13. E.g., Kelp and Simion, "What is Trustworthiness."

14. E.g., Hawley, *How to Be*, 76.

15. Kelp and Simion's version of this view cannot be interpreted along these lines since according to them, trustworthiness to *phi* merely involves being disposed to fulfill your obligation to *phi*—and not any other obligations. Kelp and Simion, "What is Trustworthiness." 3.

substantive answer—at least not without the addition of a particular view about the obligations or commitments specific to trustworthiness in discretionary circumstances.

Might we, in line with Hawley’s claim that in order to be trustworthy “you need to honor the spirit of your commitments,”<sup>16</sup> interpret these views as demanding that in discretionary circumstances we would be guided by the implicit commitments we have made or by the goals that we implicitly committed to try to aim at? As we have seen in the babysitting example above, the spirit of our commitments often underdetermines what we should do in discretionary circumstances. It may exclude some options—such as selling the baby to slavery—but it leaves many other options open. It underdetermines how to act in discretionary circumstances.

Perhaps, then, these views are best interpreted as claiming that there are no special commitments or obligations specific to discretionary circumstances, that unless, e.g., someone made a particular commitment (implicit or explicit) to act in a particular way in a specific discretionary situation, then one is free to act in any way one wishes without thereby becoming untrustworthy. If this is the correct interpretation of these views, then, as I will argue below, they are mistaken.

According to some other views, trustworthiness requires that one be motivated in a certain way. For example, Jones’ early view was that a trustworthy person must be motivated by goodwill toward the trustor.<sup>17</sup> If you have goodwill toward a person, you care about them and you are motivated by that care to promote what is good for them. But since there are diverging ideals of what is good for a person, goodwill, by itself, underdetermines one’s choices. On some views goodwill toward a person motivates you to promote their well-being, on other views, to promote the improvement of their moral character, and yet on other views, to promote their pioussness.<sup>1819</sup> Thus, while the view that trustworthiness requires

16. Hawley, *How to Be*, 10.

17. Jones\*, “Trust as an Affective,” 4, 6. Baier, in contrast, has a more complex account of the motives necessary for trustworthiness. On her view, no particular type of motive is necessary as long as the motives to keep trust override the motives to breach it. Baier\*, “Trust and Antitrust,” 253–7.

18. It is in part the open-endedness of the notion of goodwill that prompted Jones to reject her own earlier view. See Jones, “Trustworthiness.”

19. I will consider some of these ideals as guidelines for discretionary trustworthiness later in the paper.



being motivated by goodwill may exclude actions that harm the trustor, it fails to give sufficient guidance in discretionary circumstances.

According to another motive-based view, trustworthiness requires that you be motivated by a moral obligation.<sup>20</sup> Such a view might be interpreted as being implicitly committed to the claim that in discretionary circumstances trustworthiness requires that you select the most moral course of action.<sup>21</sup> I will argue against this claim later.

According to Hardin, trustworthiness requires that you be motivated by some desire or other to maintain a relationship with the trustor, and that in light of that desire you encapsulate the interests<sup>22</sup> of the trustor in your own interests.<sup>23</sup> Thus, on his view, trustworthiness requires that you be motivated so as to promote the trustor's interests. As we will shortly see, my own suggestion aligns well with Hardin's general idea that trustworthiness should sustain the relationship between trustor and trustee. But as it stands, his view underspecifies what must be done in discretionary circumstances.<sup>24</sup> Consider again the babysitting example. Prior to an analysis of the notion of "best interest," Hardin's rule—promote the best interest of the trustor—underspecifies which actions I should take to be trustworthy. Should I bring her with me to London? Should I leave her with the new neighbor?<sup>25</sup>

Finally, some views of trustworthiness speak not of the motives of the trustee but rather of the considerations to which they must be responding. For example, according to Faulkner,<sup>26</sup> a trustworthy person will be responding to the fact that they are being trusted. Like the first family of

20. McLeod, "Trust."

21. It may be interpreted this way, but it need not be. The view, as well as other motive-based views, is also compatible with the claim that trustworthiness makes no requirements in discretionary circumstances. *Morality*, of course, requires that you behave morally in discretionary circumstances, but that is a different matter.

22. Or at least enough of these to keep the relationship going.

23. Hardin, *Trust & Trustworthiness*, 1, 5, 28.

24. If "act on best interest" is to be interpreted as "act so as to maximize well-being" then Hardin's view aligns with the fourth principle I consider below. I argue against this principle after introducing it.

25. Must the notion of "best interest" be specified for it to play a role in our analysis of trustworthiness? Can we not say that the trustee ought to be guided by the best interest of the trustor and leave it to others to say what "best interest" means. The arguments I present below show that unless we specify that notion in one particular way, we fail to grasp the nature of trustworthiness.

26. Paul Faulkner, "The Practical Rationality of Trust," *Synthese* 191 (2014): 1978.

views which I have mentioned above, Faulkner's view underspecifies what this fact requires in discretionary circumstances and is thus neutral between the five positions I consider in this paper.

In the next sub-sections, I evaluate the first four suggestions for a principle of choice in discretionary circumstances. I conclude that all four must be rejected.

#### V. THE FIRST PRINCIPLE: ANY ACTION IS AS TRUSTWORTHY AS ANY OTHER ACTION

The first principle is strictly speaking not a principle but rather the absence of a principle. The rationale behind it is as follows: what you have been entrusted to do cannot be done. There is nothing else you have promised your trustor, and you are thus not bound by any promise. You are therefore free to do whatever else suits your fancy and in particular follow any guidelines you like. As just presented the principle is a non-starter. If you agreed to babysit a baby and the parents are not back on time you certainly cannot just leave the baby by himself. However, if we interpret the principle as bound in part by your implicit commitments it might well be a reasonable contender. Nevertheless, we should reject it.

Imagine Deedee is asked by Limor to deliver a package to the Clarkes and on her way learns that the Clarkes have died. According to the first principle, it is okay for Deedee to throw the package to the garbage and go to the movies, if that is what she feels like doing.<sup>27</sup> But that does not seem like a trustworthy thing to do.<sup>28</sup>

27. To be clear, it would be okay even if she did not feel like doing that. The first principle does not demand of the trustee that she follow her desires; it just says that any action is okay, including actions that are guided by one's desires.

28. Here and elsewhere in the paper, I make my argument by considering cases of entrusting. This may give the false impression that I presuppose that the fundamental relation of trustworthiness is a three-place relation (A is trustworthy to do C when entrusted by B) as opposed to a two-place relation (A is trustworthy toward B). That is not the case! For this debate see: Baier, "Trust and Antitrust," 236; Jacopo Domenicucci, and Richard Holton, "Trust as a Two-place Relation" In *The Philosophy of Trust*, ed. Paul Faulkner and Thomas Simpson (Oxford University Press, 2017); Paul Faulkner, "The Attitude of Trust is Basic," *Analysis* 75, no. 3 (2015): 424–9; Richard Holton, "Deciding to Trust Coming to Believe," *Australasian Journal of Philosophy* 72 (1994): 67; Jones, "Trust as an Affective," 6; Karen Jones, "Trust, distrust, and Affective Looping," *Philosophical Studies* 176 (2019): 955–68; Kelp and Simion, "What is Trustworthiness," 2.

A proponent of the first principle might reply that Deedee in fact had an implicit commitment to send the package back to Limor. After all, every postal and delivery service in the world makes the explicit commitment to either deliver the package or deliver it back to the sender. Such a well-known institutional commitment must also determine what is implicitly required of Deedee. If such an implicit requirement existed, then the situation would not really be discretionary, and our dissatisfaction with Deedee's choice would have to do with her failing to act as she, implicitly, committed to act, and not with a faulty choice in discretionary circumstances.

Consider, however, the following variation. Limor lives in Tokyo. Deedee, her good friend, lives in Jerusalem. Limor entrusts Deedee to purchase an expensive bouquet of fresh flowers and to deliver it to Samir who also lives in Jerusalem. After purchasing the bouquet and on her way to Samir's house, Deedee learns of his untimely death. She could, in line with the purported implicit commitment, send the bouquet to Tokyo, but there is no point since by the time it will get there the flowers will have completely wilted. She could try to return it to the store but let us assume that the florist sold Deedee his last bouquet and went out of business. Deedee throws the flowers to the garbage and goes to see a movie, because, as stipulated, that is what she feels like doing. Limor will surely think that this was an untrustworthy action. While we don't yet have an account of what the trustworthy action is, we can certainly imagine many options that would be more trustworthy. Deedee can try to deliver the bouquet to the Samir's children, she can give it to charity, or she can even keep it for herself. Any of these options would be more trustworthy than throwing the flowers away.<sup>29</sup>

We can conclude that it is not the case that the implicit rules determined by the combination of contextual factors and known institutional arrangements fully specify what to do in all discretionary circumstances. With this objection out of the way, we can reject principle number one.

29. I will have more to say in later sections about whether there is indeed a range of trustworthy actions in discretionary circumstances. For now, my point is merely that our pre-philosophical intuitions recognize a range of actions as trustworthy and see Deedee's action as untrustworthy.

VI. THE SECOND PRINCIPLE: TRY TO COMPLETE THE TASK ENTRUSTED TO YOU AS BEST YOU CAN

The idea of implicit commitments may give rise to the current suggestion. After all, isn't a request to complete a task also the implicit request to try to complete it to the greatest extent possible?<sup>30</sup>

There are two reasons to reject this principle. The first is that, as we have seen above, it is often impossible to make any progress toward completing the task entrusted to you. You cannot partly deliver a package to a dead person but you must still decide what to do with the undelivered package. The second reason is that even when it is possible to partly complete a task, it is often not the trustworthy thing to do. Your boss asks you to buy four fancy pastries for four important guests. You go to the only bakery in town and there are only three pastries left. Knowing the guests, you realize that bringing pastries for only three people will offend one of them. It is more trustworthy not to bring any. The second principle should be rejected.

VII. THE THIRD PRINCIPLE: ACT AS MORALITY REQUIRES OF YOU

I suspect that what motivates this principle is the thought that the trustee, in deciding to act in a trustworthy way, follows the demands of morality. If this is so, one might argue, then the thing to do when she cannot complete the task she was entrusted with, is to follow the remaining demands of morality.

The third principle should also be rejected. In *The Godfather*, Michael Corleone asks Al Neri to kill Don Barzini in front of the supreme courthouse. Imagine that the appearance of a police car prevented him from doing so. The demands of morality are to forgo the murder and reveal the plan to the cops. This would clearly not be the trustworthy thing to do.<sup>31</sup>

30. Thanks to Ariana Shemmer for making this suggestion, and for pushing me to clarify why the implicit instructions given at the time of entrusting could not specify what must be done in discretionary circumstances.

31. Others have made similar arguments. See Amy Mullin, "Trust, Social Norms, and Motherhood," *Journal of Social Philosophy* 36 (2005): 319. <https://doi.org/10.1111/j.1467-9833.2005.00278.x>. Note that I do not assume that revealing the plan to the cops is untrustworthy because Corleone would feel betrayed. Someone may mistakenly feel betrayed when no trust has been broken. Rather the intuition driving this example is the very mundane one that there is such a thing as trust among criminals.

This example also tells against the view that to be trustworthy in discretionary circumstances is to fulfill, or be disposed to fulfill, one's obligations. Such a view entails a variant of the third answer, where moral obligations are replaced with obligations simpliciter. But if we set up the case in such a way that Al Neri has no obligations other than his moral ones, we end up with the same conclusion. Fulfilling his obligations wouldn't be the trustworthy thing to do.<sup>32</sup>

An independent objection to the third principle emerges when we consider the possibility that several options in discretionary circumstances would be equally moral. Proponents of the principle now face a dilemma. Either they claim that a further principle of action, which by assumption goes beyond morality, is needed in these situations, and thus concede that the third principle cannot be a complete principle of action for trustworthiness in discretionary circumstance. Or they claim that all moral options are equally trustworthy in discretionary circumstances. But this claim conflicts with our intuitions. For example, it conflicts with intuitions about our opening babysitting case, where it seems obvious that the preferences of the parents (were they to be known) should play some role in choosing between the equally moral options.

It may help, in clarifying my objection to the third principle, to emphasize that the task of this paper is to investigate the standards internal, or specific, to the concept of trustworthiness. By "internal/specific" standards I do not mean necessarily non-moral standards. Internal/specific standards can be either non-moral or they can be moral, yet still independent of the totality of moral demands; for example, independent of the demands made by other virtues or by other moral considerations. It is a separate question whether being trustworthy is important according to, or required by, standards external to trustworthiness. And it is equally a separate question, assuming that the demands of trustworthiness *were* moral, how they would combine with other moral demands. These are questions I do not address in this paper.

32. Proponents of the view that trustworthiness is nothing more than a disposition to fulfill one's obligations (call them "obligationists"), might argue that the counterexample begs the question against their view since it assumes that all obligations are independent of the act of entrusting. But an obligationist who concedes that some obligations emerge with the act of entrusting must specify what these obligations are. Until he does his view is incomplete. The resulting lacuna is precisely the one that the current paper aims to fill.

A proponent of the third principle might insist that the ordinary choices of trustworthy agents, and the ordinary expectations of trustors, attest that this is the right principle after all. When I ask you to mail a package for me as fast as possible, I expect you not to try to expedite things by killing the two people standing ahead of you in line at the post office. Even mafiosi expect each other to follow the internal “moral” code adhered to by the mafia.

The proponent of the third principle has a point. A trustee is often implicitly expected to follow the moral standards of the society they are part of. But first, as the mafia case shows, these are not necessarily the standards of morality. And second, these expectations are only a working default. If I ask you to do something immoral and you agree, then the expectation that you follow societal moral standards has been canceled. Following the demands of morality is therefore not necessary for trustworthy action. Given that fact, it would be very odd, to say the least, if morality were the guiding principle of discretionary trustworthiness. What is nevertheless true is that, unless we have evidence to the contrary, it is a reasonable interpretive presumption that trustors would want you to act morally.<sup>33</sup> As we shall see later, this kernel of “truth” in the third principle is in fact better captured by the fifth principle.

A final objection applies to the third and to the first principle (“In discretionary circumstances any action is as trustworthy as any other”). According to both of these principles, in an important range of circumstances, there is an unacceptable disconnect between trustworthiness on the one hand and the personal connection between trustor and trustee on the other. According to both, in discretionary circumstances the trustworthy choice is independent of the trustor’s request, the trustor-trustee relationship, or any other relational attitude of the trustor toward the trustee, expressed or implicit. Such a disconnect is unacceptable because trustworthiness is by its very nature the

33. For a similar view of the role of interpretation in delineating the duties of fiduciaries see Joshua Getzler, “Ascribing and Limiting Fiduciary Obligations: Understanding the Operation of Consent” In *Philosophical Foundations of Fiduciary Law*, ed. Andrew S. Gold and Paul B. Miller (Oxford University Press, 2014), 49.

appropriate reaction of a trustee to their personal<sup>34</sup> relationship with a trustor.

#### VIII. THE FOURTH PRINCIPLE: ACT SO AS TO MAXIMIZE THE WELL-BEING OF THE TRUSTOR

The fourth principle might be motivated by the combination of two thoughts. First, that to be trustworthy one must be moved by goodwill toward the trustor. And second, that the best way to understand goodwill toward a person is as a concern for their well-being.

This principle should also be rejected. Deedee and Alfie go shopping for Limor who was diagnosed with Covid 19 and is now self-isolating. Limor says: "Please get me a 3% fat yogurt." When they reach the store all they can find are fat-free and 5%-fat yogurts. They try to call Limor, but there is no reception. Deedee puts a fat-free yogurt in the cart. "She doesn't like fat-free yogurts," says Alfie, "she would prefer the 5% one." Deedee replies: "It doesn't matter. Low fat is better for her."

Importantly, this last counterexample does not depend on the thought that Deedee has a mistaken view of what is required for Limor's well-being. The counterexample works even if fat-free yogurt is objectively best for Limor.

I find the intuition that Deedee is not trustworthy rather strong. But what can I say to someone who does not share that intuition? One thing to say is that I assume the case to be one where Deedee selects a yogurt that she knows Limor would not be happy to receive. The case would be rather different if Deedee anticipated that Limor would be happy with Deedee's choice; perhaps, for example, because Limor would see it as a welcome attempt to help her (Limor) combat her weak will. I discuss such cases later in the paper.

A different set of examples might be said to support the fourth principle. Fiduciaries, medical workers and other carers, who often need to

34. Annette Baier says that when we enter a library, we trust our fellow library dwellers to push us out of the way were they to see a stack of books on an upper shelf about to fall onto our head. Baier, "Trust and Antitrust," 237. This trust is independent of any personal relationship we may have with them. Couldn't the relationship which grounds the duties of trustworthiness be the relationship between every human and their fellow humans? And aren't, therefore, the demands of trustworthiness the same as the demands of morality after all? Baier may be right about library dwellers, but the conclusion overgeneralizes. If you are in the midst of an armed robbery of the old manuscript collection, I may well think that your failure to prevent a box from falling on my head is immoral, but I will not think it is a failure of trustworthiness.

make decisions in discretionary circumstances, are often expected to consider the well-being of those under their care. This expectation finds written expression in texts as early as the Hippocratic Oath<sup>35</sup> and the First book of Plato's Republic,<sup>36</sup> where it appears as the requirement to benefit (or be guided by the "advantage of") those under your care. The very same expectation is expressed in both contemporary fiduciary law<sup>37</sup> and in many contemporary mental capacity acts<sup>38</sup> as the requirement to be guided by the best interest of one's beneficiaries or those for whom one makes decisions. But one shouldn't confuse trustors and beneficiaries.<sup>39</sup> The duty of trustees, according to fiduciary law, is to be guided by the best interest of their beneficiaries<sup>40</sup> not the best interests of those who have set up the trust. Similarly, the duty of those making decisions for people who lack mental capacity is to be guided by the best interest of those on behalf of whom the decision is being made and not those (judges, society, or family members) who have entrusted them with that task. The nature of that duty is unsurprising. If you take it upon yourself to care for someone then you should be guided by what is good for them. But these are all cases where the trustee has been entrusted, specifically, to act in a way that is guided by the interest of someone. In other words, these are all cases where the trustor set up the trust relation precisely because they wanted the trustee to be guided by the interests of a third-party. Indeed, in most of these cases, it is an implicit instruction of the legal system, and in some legal contexts an explicit instruction, that once the trust is in place, one should ignore the interests of the trustor when deciding how to best benefit the beneficiary. Note also those cases where an organism might be entrusted by the government to a biologist precisely with the aim of destroying it—maybe because leaving it alive endangers public

35. Ludwig Edelstein, *The Hippocratic Oath: Text, Translation and Interpretation* (Johns Hopkins Press, 1943).

36. C.D.C. Reeve, *Plato, The Republic* (Hackett, 2004), 342e.

37. E.g., Trustee Act 2000, c. 29, accessed June 25, 2024, part IV, section 15, 3. <https://www.legislation.gov.uk/ukpga/2000/29/contents/enacted>.

38. E.g., *Mental Capacity Act 2005*, c. 9, accessed June 25, 2024, part 1, Section 1, principle 5. <https://www.legislation.gov.uk/ukpga/2005/9/contents/enacted>.

39. Trusts are often set by one party for the benefit of another. A dying parent might set up a trust to be managed after their death by a trustee for the benefit of their children. The parent is the trustor, but the children are the beneficiaries. The duty of the trustee is to the children not to the parent.

40. Leonard I. Rotman, "Fiduciary Law's 'Holy Grail': Reconciling Theory and Practice in Fiduciary Jurisprudence," *Boston University Law Review* 91, no. 3 (2011): 940.



health. In those cases, the guiding principle of the biologist should be the worst, and not the best, interest of the organism entrusted to them.

The above discussion of fiduciaries and carers reminds us of three things. First, beneficiaries are not identical with trustors. It thus does not follow from the fact that a trustee ought to be guided by the best interests of beneficiaries, that she should be guided by the best interests of her trustors. Second, while in many cases the trustee is expected to care for those people or these things entrusted to them, this is not always the case. Third, that often, when a trustee is a fiduciary, the best interest of the trustor is explicitly stated not to be the guiding principle of their action. As we will see below, the particular cases of a trustee who is entrusted with the best interest of a beneficiary, or of a care worker who is entrusted with making decisions on behalf of a person lacking mental capacity, are also better captured by the fifth principle.

In the next section, I will claim that the fifth principle is the best candidate for a standard of trustworthiness in discretionary circumstances. It emerges naturally once we consider the common problem with the previous principles. They all, one way or another, failed to respect the would be wishes of the trustor. As I will argue later, the fifth principle also captures the “kernels of truth” in these other principles.

#### IX. SECTION FOUR

We have considered four principles and found them unsatisfying. We are left with the fifth. Let that principle serve as an initial version of the correct standard of trustworthiness in discretionary circumstances:

*Act as the Trustor would have Wanted you to Act, in order to achieve his/her aim,<sup>41</sup> had she deliberated about the discretionary circumstances.<sup>42,43</sup>*

41. The qualification within the commas ought to be clarified. Trustworthiness with respect to an entrusted task in discretionary circumstances might require varying one’s actions in order to achieve the original aims of the trustor. One is not required, in such circumstances, to try to fulfill other aims of the trustor. Thanks to Chris Bennett for convincing me of the necessity of this qualification.

42. There may be more than one thing that the trustor would be equally happy for you to do. In that case you can select any one of those things, and you are free to let other considerations impinge on that selection. Thanks to Sarah Durling for noticing this possibility.

43. The principle, as stated, asks what the trustor would have wanted at the time of entrusting. A gap could open between what they would have wanted at the time of entrusting (that you do in the discretionary circumstances) and what they would have wanted at the time of the discretionary circumstances. I leave for a different time the discussion of how such gap should be handled. Thanks to Ilya Shemmer for pushing me to clarify this point.

Call this principle TWA<sup>initial</sup>.<sup>44</sup> In the next two subsections, I consider objections to TWA<sup>initial</sup> and refine it in light of these objections.

#### X. THE “TOO ONEROUS” OBJECTION

There are numerous cases in which “it would be crazy” to do what the trustor would want you to do in the discretionary circumstances.<sup>45</sup> The trustor might want you to do something utterly immoral, disgusting, embarrassing, or humiliating. The trustor might want you to do something which would require that you expend inordinate financial resources, spend an inordinate amount of time, or exert an inordinate physical effort.

“*It would be crazy* to do what the trustor would want you to do” is a vague claim. We can be more precise. These are all cases in which the action required by TWA<sup>46</sup> is too onerous. One may be willing to breach certain moral principles up to a point to be trustworthy.<sup>47</sup> Beyond that point, the breach of principles involves too great a moral compromise. One may be willing to do something hurtful to a person they care about in order to be trustworthy, but only up to a point. Beyond that point, hurting someone you care about is going to be too immoral or too painful for you. One may be willing to spend quite a bit of money to be trustworthy, but only up to a point. Beyond that point, the expenditure is too large. We can

44. I have mentioned above that according to the Mental Capacity Act 2005 UK, people who make decisions on behalf of another are meant to be guided by the best interest of the person for whom the decision is being made. The situation is more complicated in the United States. Different jurisdictions are governed by different codes. One typical opinion is that of the American Medical Association. Its code of practice requires that the principle of Substituted Judgment (which doctors often understand as grounded in the ideal of autonomy) be the primary guide in decisions made for another person. According to that principle, a decision maker is meant to guide herself by her best estimation of the wishes of the person for whom she decides. One might think that this last approach supports the fifth principle. But this isn't necessarily so. First, the relevant relation between patient and decision maker is often not understood as a relation of trust. Second, and relatedly, the principle of Substituted Judgment is often understood as governing the *moral* obligations of carers and decisions makers and not their obligations qua trustees.

45. Thanks to Ruth Weintraub for suggesting and discussing this objection.

46. Since the objection applies to later versions of TWA as well, I drop in the remainder of this subsection the superscript “<sup>initial</sup>.”

47. I should reiterate that my discussion only concerns itself with what is allowed/required by the standards internal to trustworthiness. Morality, of course, condemns all immoral actions.

understand all these cases as situations in which the counterfactual desire of the trustor would demand too much of the trustee.

TWA makes no exceptions for cases in which acting on the counterfactual desires of the trustor would be too onerous. That seems wrong.

Our next question is, therefore, whether we can analyze what it is to be “too onerous,” and whether we can use this analysis to modify TWA so that it meshes with our intuitions about which actions keep trust in those circumstances.

To be sure, we are not interested in the notion of being “too onerous” in general. It may well be that in other contexts, e.g., when we are asked for a favor, “being too onerous” sets up different limits on what we ought to agree to do. Our interest is only in what it is to be “too onerous” in the context of trustworthiness in discretionary circumstances. I therefore suggest that instead of trying to figure out what “being too onerous” means, we keep this notion as an accompanying metaphor and focus directly on the bounds we can impose on what it would be too much to expect of a trustee in discretionary circumstances.

We can set an Upper Bound (UB) on what one must do in discretionary circumstances, in order to count as trustworthy, by appeal to a counterfactual. UB\_1: *If at the time of entrusting the trustee would have refused to perform a certain action A, if it were asked of them, and the trustee now, in the discretionary situation, refuses to perform A, then the current refusal does not count against their trustworthiness.*

In other words, and on the assumption that the surrounding circumstances have not changed from the time of entrusting, a trustee would not be less trustworthy if they now refuse to perform actions that they would have refused to perform if the actions were asked of them explicitly, with the discretionary circumstances foreseen, at the time of entrusting.

Let us take an example: Limor entrusts Deedee with the task of delivering a package to a destination which is 3 miles from Deedee’s point of origin. Deedee agrees. En route, Deedee discovers that Limor was, through no fault of her own, wrong, and the delivery destination is in fact 40 miles away. She refuses to drive that far. We can now apply our counterfactual test. If Limor had asked her to drive 40 miles to start with and she would have said no, then 40 miles now is too onerous, and Deedee’s current refusal does not detract from her trustworthiness.

Consider also an example from the moral domain. Don entrusts Al the petty criminal with the task of collecting protection money from the local grocery store. Al has been happily collecting protection money for years. However, today the shopkeeper refuses to pay. Al finds himself in discretionary circumstances. He knows that Don would have liked him to break the shopkeeper's nose in order to make him pay. But Al has his limits. He is happy to be "in crime" as long as no-one gets physically hurt—physically hurting innocent people, thinks Al, is something that morality prohibits him from doing. If Don would have asked Al at the time of entrusting to break the shopkeeper's nose, he would have refused.<sup>48</sup> Therefore, according to UB\_1, Al does not fail to keep trust if he does not break the nose of the shopkeeper.

It might be objected that there is a less convoluted way to explain why the demand to break someone's nose in discretionary circumstances is too onerous. It is too onerous, says the objector, because it is immoral. Maybe trust requires that you act immorally if you promised to do so, and maybe it doesn't, but it certainly cannot require, says the objector, that one act immorally if one has not promised to do so.

I have already argued that morality should not serve as the overall determining principle in discretionary circumstances. We can think of the current objection as an attempt to partially push back against my earlier argument by insisting that even if morality is not the overall principle guiding action in discretionary circumstances, it should certainly be a constraint on what one should do in those circumstances. In support of that view, the objector puts forth the claim that an immoral action is always too onerous if one has not agreed to be entrusted with it.

The objection should be rejected. Consider a variant of our last story. Unlike Al the petty criminal who objects to violence, Arnold the henchman delights in it. He has been Don's number one "nose breaker" for many years, and he gets great pleasure from doing so. Now, as before, the shopkeeper refuses to pay, and Arnold realizes that Don would have wanted the shopkeeper's nose broken. It makes no sense to insist that breaking one more nose is too onerous for Arnold, certainly not on the grounds that such action is immoral.

48. Let us imagine that Al is a childhood friend and is therefore in a unique position to refuse Don's requests.

One might argue that UB\_1 is of very little practical use to anyone but the trustee<sup>49</sup> since she is the only one who has access to her own counterfactual commitments. I am not sure that this is true (see more below), but even if it were true, it wouldn't detract from the claim that this principle designates correct bounds on trustworthy action in discretionary circumstances. To see this, it suffices to notice that the question of whether trust was kept is a question that a person can ask about their own actions. If UB\_1 helps us answer that question about ourselves, in a moment of honest self-reflection, then it can be used to delineate the conceptual boundaries of trustworthiness even if they are of little practical use to a third-party. Surprisingly, and in a different way, the epistemic difficulties that accompany UB\_1 help vindicate our analysis. I return to that point, as well, below.

The problem with UB\_1 is that it does not provide for cases where the surrounding circumstances, other than the discretionary ones, have significantly changed. We need an upper bound that addresses these cases too. This is particularly true when the trustee herself has undergone significant changes. Deedee is entrusted by Limor to deliver a package by car. If asked at the time of entrusting, Deedee would have agreed to walk a further mile to complete the delivery. After being entrusted with the delivery and before the discretionary situation arose, Deedee breaks her left leg. Having driven to her intended destination (in an automatic car), Deedee discovers that the recipient lives on a rural farm and that the only way to the house requires walking an extra mile. Suppose that Deedee knows that Limor would have wanted her to walk the extra mile, and suppose she decides not to walk. It is obvious, says the objector, that her current refusal to walk in order to complete the delivery should not count against her trustworthiness.

That is a good point, but we can accommodate it with a minor amendment. We can ask whether the trustee would have refused to perform a given action if it were asked of them at the time of entrusting *on the assumption that the trustee already knew about the (forthcoming) change*

49. I do not assume that it is always transparent to the agent what she would have agreed to in counterfactual circumstances, and I certainly don't deny that there are borderline cases in which it would be very hard to know how you would have reacted to a trustor's request. However, all I need in order to show that UB\_1 correctly characterizes trustworthy action in discretionary circumstances is that on many occasions agents would have a good sense of what they would have agreed to in the counterfactual moment of entrusting.

*in circumstances and fully appreciated their impact.* Call the amended principle UB<sub>2</sub>.<sup>50</sup> Thus, we can ask whether Deedee would have refused to walk one mile had she already known that when the time came to deliver the package her leg would be broken, and had she fully understood the impact that a broken leg would have on her. If at the time of entrusting Deedee would have refused to walk a mile on a broken leg then, on the amended principle, her current refusal would not count against her trustworthiness.<sup>51</sup>

A further worry about UB<sub>1&2</sub> concerns situations where the trustee's answer to the question embedded in the counterfactual test is arbitrary or irrational.<sup>52</sup> Imagine that if asked, Deedee would have agreed to carry 4 pounds of self-rising flour for Limor, or 4 pounds of oranges, or 4 pounds of nuts, but would have refused to carry 4 pounds of regular flour. Imagine further that in the discretionary situation she would have needed to carry 4 pounds of regular flour to satisfy what Limor would have wanted her to do, and finally imagine that she in fact refuses to do so. According to UB<sub>1&2</sub>, the match between her current refusal and her counterfactual refusal at the time of entrusting should prevent us from counting her as untrustworthy. An objector might argue that this is a contentious result.

The objection can be understood in two different ways. On the first understanding, the objector claims that certain responses of the trustee to the question in the counterfactual test are unacceptable because they clash with an objective standard about what a person in the circumstances and with the abilities of the trustee should be willing to do for the trustor. For example, the objector might insist that someone as strong as Deedee, who, like Deedee, has nothing better to do with her afternoon, should agree to carry 4 pounds of regular flour for Limor. If this were the ground

50. I present the final version in full after I introduce one more amendment.

51. UB<sub>1&2</sub> apply most naturally to those cases where in the counterfactual situation in which the trustee would have refused to perform certain actions in discretionary circumstances, the trustor was, counterfactually, nevertheless happy to entrust her with the task in question. But what should we say about those cases in which a counterfactual refusal to fulfill a potential desire of the trustor would have prompted the trustor, had she known about it, not to entrust? There are two different scenarios to be considered. If it should have been obvious to the trustee that his refusal would have annulled the act of entrusting, then there is more pressure on the trustee, though not a requirement, to satisfy the counterfactual desire of the trustor even if that exceeds UB<sub>1&2</sub>. On the other hand, if there was little reason to think that the trustee's refusal would have annulled the act of entrusting, then there isn't any pressure on the trustee to satisfy the counterfactual desire of the trustor.

52. Thanks to Ilya Shemmer for raising and discussing this objection.

of the objection, then I think we should reject it. Indeed, the counterfactual test set up by UB\_1&2 is designed precisely to capture the subjective aspect of our understanding of being “too onerous.” Deedee might have a longstanding personal animosity to regular flour. Maybe a bag of regular flour was responsible for a significant and traumatizing childhood misfortune. It might be very important to Deedee, but not to others with her abilities and in her circumstances, never to transport this type of flour. For Deedee, carrying 4 pounds of regular flour *is* too onerous, though it may not be too onerous for others. The fact that UB\_1&2 capture this subjectivity is a feature, not a bug.

On a second understanding of the objection, the objector accepts our insistence that what counts as “too onerous” depends crucially on the desires, concerns, goals, and cares of the trustee, but worries that the trustee’s response to the question embedded in the counterfactual test may be arbitrary or irrational relative to these very desires, concerns, goals, and cares. Thus, it makes perfect sense for the trustee to refuse to transport meat if she is a vegetarian, but it makes no sense for the trustee to answer the question of the counterfactual test by picking an action arbitrarily or irrationally, given her own desires. The fact that such an arbitrary/irrational would-be answer matches her future refusal is no defense against the judgment that her behavior is untrustworthy.

Understood in this second way, the objection is, I believe, correct. To deal with it we need to introduce a second amendment to our counterfactual test (UB\_2). The question we should be asking is not whether the trustee would have in fact refused, if asked at the time of entrusting, to do what the trustor would have wanted done in the discretionary circumstances, but rather whether the trustee *would have refused were she to answer non-arbitrarily and rationally in light of her desires, concerns, goals and cares*. To simplify the phrasing, I will refer to this counterfactual refusal as “rational refusal.”<sup>53</sup>

Our objector might not be satisfied. The current amendment, she might claim, addresses the worry that a trustee might make arbitrary or irrational choices given their desires, but it does not address the worry that her desires

53. As Ilya Shemmer correctly observed, even if we accept the claim that in assessing what is onerous, we should only consider rationality and arbitrariness relative to one’s cares and concerns, there is a sense of “arbitrary” in which what a person would counterfactually agree to in order to do what they were entrusted with, is unavoidably arbitrary. Instead of agreeing to stay somewhere for 2 h the trustee could have agreed to stay for 2 h and 3 s. Such arbitrariness does not undermine trustworthiness.

themselves might be irrational. If the trustee was traumatized by regular flour as a child, then maybe carrying this type of flour is truly hard for them, but not so if they have an aversion to carrying regular flour because a talking fly convinced them that it is a bad idea to carry regular flour.

I am only partly convinced by this objection. Our desires, cares, concerns and aversions are real regardless of their origins. My fear of spiders, albeit irrational, is real. Walking through a room full of spiders is much harder for me than it would be for someone who is not possessed by that fear. A person's preference of Christ over Allah may be irrational, but it is nonetheless true that it would be much harder for them to denounce the one than to denounce the other. Grounding UB in real desires, concerns, aversions, and fears, regardless of their origin, makes perfect sense.<sup>54</sup> This last claim is not meant to betray a general preference for subjectivism in metaethics. It is merely the claim that in assessing how hard a certain task is for a person we need to take into account their own, possibly idiosyncratic, preferences.<sup>55</sup> I do however acknowledge that there is a certain type of irrational desire that cannot serve as the basis of UB: these are irrational desires that are specific to the trustor or to the group that the trustor belongs to. If Deedee has an irrational aversion to doing things that Limor wants her to do or has an irrational aversion to doing things that people who live in Japan want her to do, then that aversion cannot exempt her from judgments of untrustworthiness. Entering a relation of trust while harboring such discriminatory attitudes toward the trustor amounts to entering it in bad faith; bad faith which explains why we see the counterfactual refusal to act as the trustor would have wanted as untrustworthy. Let us call refusal that is rationally grounded in one's desires and aversions, but which excludes discriminatory attitudes toward the trustor "rational and non-discriminatory refusal."

Incorporating this last amendment, we get UB final (henceforth UB):

54. On many analyses, trustworthiness has both a motivational component and a competence component, e.g., Hawley, "Trust, Distrust." Irrational aversions may well reduce your competence to perform certain actions and thus affect your overall "trustworthiness score." It would be interesting in the future to consider the way this complication affects our analysis of the competence component of trustworthiness. That fact is, however, unrelated to the analysis of what is truly too onerous for you.

55. It is a separate question whether the irrationality of one's desire, by itself, undermines trustworthiness. I cannot see why it would and I therefore leave it to others to argue one way or another.



*UB: If at the time of entrusting the trustee would have, rationally and without discriminating against the trustor, refused to perform a certain action A, on the assumption that she already knew about the changes in circumstances surrounding the discretionary situation and fully appreciated their impact, and the trustee now, in the discretionary situation, refuses to perform A, then the current refusal does not count against their trustworthiness.*<sup>56</sup>

It is worth emphasizing that further thinking might help us locate a lower upper-bound. I am not committed to the claim that UB is the best we can do. Identifying it merely shows that the vague idea of being “too onerous” can be given an exact expression in light of our intuitions about trustworthiness, and that it, therefore, can be combined with TWA to give us a principle that determines trustworthiness in discretionary circumstances.

This concludes my discussion of the “too onerous” objection.

#### XI. THE “NON-IDEAL TRUSTOR OBJECTION”

If the trustor were asked what they would want you to do in a discretionary situation, they might give an answer that does not make sense given their (the trustor’s) own interests, goals and principles. Why? For the same reasons that people sometimes wish for things that are not good for them. Maybe they are afraid, maybe they are angry, maybe a desire of theirs has a disproportionate motivational influence on their decision making, etc.

TWA<sup>initial</sup> asserts that one must respect the counterfactual wishes of the trustor in discretionary circumstances (and we can now add: within the bounds imposed by UB) in order to keep trust. But at least in those cases in which the trustor does not reason well from their own interests, goals and principles, this seems wrong.<sup>57</sup>

Dora arrives in the morning at her friend Frieda’s house to help organize for a party that night. They spend the day cleaning and decorating. Dora also spends the day filling up on alcohol. When the party starts Dora

56. This second amendment further highlights the epistemic difficulties involved in using UB in deciding trustworthiness on a particular occasion. I discuss these difficulties in the last section of the paper.

57. Thanks to Graham Bex-Priestley and Novenka Bex-Priestley for raising and discussing this objection.

is already drunk. She notices her car keys on the table and gives them to Frieda for safekeeping until the party is over. She tells Frieda that she will now stop drinking so she can drive home after the party. But Dora fails to stick to her plan. When the party ends, Dora is even more drunk than she was at the start. She asks Frieda for her keys back. Frieda finds herself in a discretionary situation. She was entrusted with keeping the keys and giving them back at the end of the party. She assumed (and so did Dora) that when the party ended Dora would be sober. But as it turns out, Dora is drunk. Must Frieda give the keys back to count as trustworthy?

The answer given by TWA<sup>initial</sup> is that Frieda must give the keys back if this is what Dora would have wanted her to do. Recall that Dora was already drunk when she entrusted the keys to Frieda. Let us assume that drunk Dora is reckless and when considering what she would want done in that discretionary situation, would want to be given the keys back. In that case, TWA<sup>initial</sup> requires that Frieda returns the keys to Dora. TWA<sup>initial</sup> gives us the wrong result.<sup>58</sup>

This objection points toward a natural solution. We could modify TWA<sup>initial</sup> and ask not what the trustor would have wanted in the discretionary situation but rather what an idealized trustor would have wanted in this situation. An idealized Dora would not have wanted Frieda to give her keys back. An idealized Dora would have realized that driving under the influence is not a good idea and would not have let her drunkenness affect her preferences.

But idealized how? This is a complicated question that we would have to broach were we to try to address the objection in the way suggested in the previous paragraph. Conveniently, other cases raise problems for this approach, however, we specify it.

First case against idealizing.<sup>59</sup> You ask me to drop your CV in the mailbox on Barber Road. I see it is addressed to the philosophy department at Leeds University. When I discover that the mailbox at Barber Rd was destroyed by vandals, I figure that you would have wanted me to take it to the next closest mailbox on Crookesmoor Rd. TWA<sup>initial</sup> says that this is what I should do to keep trust. But I am “smarter” than that. I reason that

58. One could try to deal with the objection by appeal to UB above. Let us assume, for the sake of argument, that giving the keys back is not “too onerous” for Frieda. As we will soon see, in other cases discussed in the context of this objection, it makes no sense to use UB as a way of addressing the problem.

59. Thanks to Novenka Bex-Priestley for suggesting this case.

your ideal self would have known that an academic career in philosophy is a poor choice.<sup>60</sup> So, instead, I put your CV in the shredder.

Second case against idealizing. Cyrano, with usual panache, writes a stylish letter, seals it, and asks his friend Edmond to deliver it to Roxane whose love Cyrano seeks. On the way there, Edmond loses the letter. He knows the content in broad outlines but has not seen the phrasing. When Roxane asks what message he had brought her from Cyrano, Edmond has to make it up on the go. He is familiar with Cyrano's flamboyant style, and is confident he could imitate it, but he believes that an ideal Cyrano would have known that his own panache must be tamed. So, Edmond delivers Cyrano's message of love in boring prose.

This last case seems even worse. In both of the last two cases, the trustee takes it upon himself to act on the goals and desires he thinks the trustor should aspire to have. In the last case, he also takes it upon himself to represent the trustor not as he is, but as he, the trustee, believes the trustor would have wanted to be.

An objector might argue that the intuition in these cases isn't against idealizing. Our intuition, goes this objection, springs solely from our dissatisfaction with the imposition by the trustee of some goals and aspirations while lacking a true understanding of what the ideal trustor would have wanted. The objector is wrong. Even if the trustee had a true insight into what the ideal trustor would have wanted for himself, that insight very often would be alien to the actual trustor. When such an insight is imposed on the actual trustor from without, it must feel like betrayal.

Some cases pull us toward an idealized version of the counterfactual condition of TWA<sup>initial</sup>. Other cases speak against idealization. Is there a better interpretation of the counterfactual test at the heart of TWA<sup>initial</sup> that would fit our intuitions in all those cases? Should we give up on TWA altogether?

I think the best solution is to supplement TWA<sup>initial</sup>. We can do that in two stages.

In the first stage, we consider a simple supplement. First, the trustee should take the desires of the actual (non-ideal) trustor at the time of entrusting as his benchmark. He should ask himself how the non-ideal trustor at the time of entrusting would have wanted to be interpreted by

60. To avoid any confusion, we can assume that your ideal self is right.

the trustee when the latter considers what she (the trustor) would have wanted in discretionary circumstances; whether she would have wanted the trustee to act on the desires of her non-ideal self or whether she would have wanted the trustee to act on the desires of her ideal self? Second, the trustee should interpret the counterfactual condition of TWA accordingly. At the time of entrusting Frieda with the keys, non-ideal Dora would probably want that, in discretionary circumstances, Frieda decides according to the wishes of an ideal Dora; not a Dora whose false beliefs would lead to her own death. Non-ideal Cyrano, on the other hand, at the time of entrusting Edmond with his letter, would probably want Edmond to represent him to Roxane—if it came to this—as he really is, and not as the modest, pale version of himself that he would adopt were he a more virtuous person.<sup>61</sup> Cyrano, at the time of entrusting, thus wants Edmond to guide himself in discretionary situations in accordance with what the actual Cyrano would have wanted in these situations.

So far, our simple suggestion is to let the counterfactual desires of the real trustor at the time of entrusting decide between an ideal and non-ideal version of her desires for the discretionary situation. But there is no reason to be satisfied with this binary choice. We can think of different degrees of idealization, different forms of idealization, and of idealization as contextually dependent on the type of discretionary circumstances at hand. Our job applicant might be happy for me to idealize when it comes to the question of the best way to deliver the letter to Leeds University, but not want me to idealize when it comes to the goals he should have in life. Most people don't want their trustee to idealize with respect to the moral judgments that they should make. But some devotees are happy for their masters/rabbis/priests/leaders to decide for them what moral judgments they should make in discretionary circumstances. Dora might want Frieda to idealize in life-and-death situations but not in less precarious situations.

In the second stage, we offer a more complex supplement to TWA<sup>initial</sup>. That supplement demands that the trustee first asks not only whether the trustor, at the time of entrusting, would have wanted us to think of their ideal/non-ideal self as determiner of what should be done in discretionary situations, but more specifically how much and in what respects they

61. I assume here that Cyrano agrees with Edmond that a more modest personality would have been preferable, but that he is unwilling to give up on his flashy self.

should be idealized, and how they should be idealized in different possible discretionary situations. We might phrase the suggested supplement more succinctly by saying that the trustee should idealize *in the way* in which, if asked at the time of entrusting, the trustor would have wanted to be idealized.

Thus, we get the final version of our principle, simply labeled TWA:

*In discretionary situations act as the trustor would have wanted you to act had she deliberated about the discretionary situation, and had she been idealized in the way in which (if asked at the time of entrusting) she would have wanted to be idealized.*

Our complete principle of action in discretionary situations is thus TWA moderated by UB.

But what should we do when the action that the trustor would have wanted is too onerous? Strictly speaking, this is left underspecified by these two principles. However, a natural extension in the spirit of these principles demands that we try to do the thing that the trustor would have most wanted and which is not too onerous. I will label this natural extension:

*TWA-UB.*

Is TWA-UB superior to the four principles I have discussed at the start of the paper? We have already seen a number of reasons for a positive answer to that question. There is, however, another reason to prefer the fifth principle. This principle easily incorporates the kernels of truths we have found in principles 2–4. Most people most often prefer a partial completion of a task to no action at all, care about morality, and care about their best interest. The goals that most people most often care about are also the goals they would, most often, want a trustee to aim at in discretionary circumstance. But “most people and most often” is not “everyone” and not “always.” TWA-UB thus both explains the common intuitions in favor of principles 2–4 and explains our intuitions about those cases where these principles should be rejected.

## XII. SECTION FIVE

The foregoing analysis raises two questions.

- (1) Assuming there are principles determining how to behave in discretionary circumstances, does that leave space for genuine discretion? Or put the other way around, assuming there is real discretion in trust, must not the analysis provided above be mistaken?
- (2) To know what is required by TWA-UB one must figure out the answer to two questions. First, what the trustor, and sometimes the idealized trustor, would have wanted in the discretionary circumstances, and second, how the trustee would have reacted at the time of entrusting if asked to be entrusted with more than what they were in fact entrusted with. One might have partial evidence that would help to answer these questions, but it is very hard to be confident that one has done so correctly. Doesn't this epistemic difficulty<sup>62</sup> indicate that TWA-UB is, after all, the wrong principle?

Before I address the first question, I should repeat a point made earlier: the standards discussed here are internal to the notion of trustworthiness. I do not, in this paper, discuss the normative force of these standards. I leave completely open such questions as whether there is a requirement to be trustworthy, whether it is important to be trustworthy, or how to balance the demands of trustworthiness with other demands of rationality or of morality. Thus, when I ask now whether the standards of trustworthiness are discretionary, I refer again only to the standards internal to the notion of trustworthiness. It may well be that from the point of view of, e.g., morality, trustworthiness itself is completely optional. If that is the case then from the point of view of morality trustworthiness, including everything demanded by its internal standards, is discretionary. However, the question I am now pursuing is not whether such external discretion exists. Rather, what I aim to explore is the question of whether from the point of view internal to the concept of trustworthiness there is a real discretion in how one acts in "discretionary" situations. My insistence that there are standards which determine how one should behave in order to count as trustworthy suggests a negative answer to this question.

One way to resist that negative answer would be to shift our focus from trustworthiness on a particular occasion to trustworthiness as a character

62. The fact that much entrusting occurs implicitly further adds to the epistemic difficulty involved in figuring out what is required by the principles of trustworthiness in discretionary circumstances.

trait. It could then be suggested that from the point of view of one's character, the standards of action in discretionary situations impose only an imperfect demand. It is possible to breach these standards once or twice or even three times and still count as a trustworthy person. This may well be so, but it does not answer the more interesting question about whether one has real discretion in a discretionary situation when it comes to the assessment of trustworthiness on a particular occasion.

Another way to resist the conclusion that there is no real discretion in discretionary circumstances would be to maintain that the standards of action in these circumstances are supererogatory.<sup>63</sup> On that view, on the one hand, one can fail these standards and still be trustworthy. And on the other hand, when someone is guided by these standards one is not merely trustworthy but rather is a particularly shiny exemplar of trustworthiness. We should reject this suggestion as shown by some of the examples given above. If, when I see a note on the door saying that the Clarkes have moved two houses down the street, I throw the package to the garbage instead of taking it to them, I do not merely become a less glorious exemplar of trustworthiness; I completely fail to act in a trustworthy way.<sup>64</sup>

To answer the first question, I propose we make a distinction between two types of discretion, which I will label "normative" and "epistemic." Normative discretion exists if there are no, or only partial, norms specifying what one ought to do in discretionary circumstances in order to qualify as trustworthy. Normative discretion does not exist if there are norms fully specifying what to do in discretionary circumstances in order to qualify as trustworthy. Epistemic discretion can be in place even without a normative one. Epistemic discretion exists when there are serious epistemic difficulties involved in figuring out what is required by the discretionary norms in particular situations. Such difficulties leave a wide range of legitimate possibilities for the trustee to choose from, since there is a wide range of legitimate beliefs about what the norms require. For example, if a discretionary norm requires that in discretionary circumstances you should do exactly what Abe Lincoln's cousin, if asked, would have

63. The discussion in this paragraph assumes that *internal* standards can be supererogatory. This may be an unacceptable assumption, but as we will see at the end of the paragraph, I reject this approach anyway. Thanks to Chris Bennett for pushing me to clarify these issues.

64. Since I only speak of the standards internal to trustworthiness, what I say here leaves open the possibility that from the point of view of morality trustworthiness is supererogatory, or that some aspects of it—such as choice in discretionary circumstances—are.

wanted you to do, then while there is a clear standard determining the right thing to do, there is a serious difficulty figuring out what that standard requires on a particular occasion. And this means that there are many legitimate beliefs one could have about what this standard requires on the current occasion.

It is my view that the discretionary nature of trust is epistemic, not normative. The lack of normative discretion explains why it is wrong to choose what to do arbitrarily in discretionary situations, and why in many cases we have clear intuitions about what is and isn't trustworthy in those situations. The existence of epistemic discretion explains why we think that the trustee has a large leeway in selecting what to do, and why we think that the trustor should show flexibility and charity in her reactive attitudes toward the choices of the trustee.

An objection presents itself. Isn't it the case that, at least sometimes, what we trust people to do is precisely to use their own judgment in discretionary situations? On those occasions, aren't we happy to accept whatever it is that they decide? Doesn't that show that, at least sometimes, the discretion of the trustee is normative and not merely epistemic?

The objection is correct in supposing that trust sometimes involves a willingness, maybe even a desire, that the trustee use their own judgment, unguided by their understanding of how the trustor would have wanted them to handle the discretionary situation. However, it is mistaken in concluding that trustworthiness involves normative discretion. Even in the situations highlighted by the objection a trustworthy trustee complies with the more general counterfactual desire of the trustor, that is, with the desire that the trustee use their own judgment about what to do. Crucially, this general desire is not a necessary component of a trusting relation.

To recap, our answer to the first question is that there can be discretion despite the existence of clear principles of action, as long as epistemic discretion is in place.

This leads directly to a threefold answer to the second question:

- Not only do the epistemic difficulties involved in applying the principles of action in discretionary situations not undermine our account, they in fact strengthen it. It's precisely those difficulties that allow us to think of the "discretionary nature of trust" as truly discretionary.
- These principles have an important theoretical role to play even if on occasion they fail to have any practical import. As I have explained above, they delineate the boundaries of the concept of trustworthiness.
- Often these principles are also useful as practical guidelines. How useful they are depends in part on how well the trustor and trustee know each



other. This insight underscores an important and hitherto undiscussed connection between trust and friendship. The intimate knowledge that friends have of each other's desires, preferences, goals, and principles, either in general, or in particular sub-domains, makes friends more likely to be better trustees and better trustors.<sup>65</sup> More generally, understanding the epistemic difficulties which face those who wish to comply with the principles of trustworthiness in discretionary circumstances helps us identify who would be a good trustee for particular trustors, and for particular occasions of entrusting.

### XIII. CONCLUSION

I have explored the internal standards of trustworthiness. I have argued that in discretionary circumstances the trustee does not have a *carte blanche* to act as she sees fit. She must be guided by the wishes that the trustor would have had concerning the action of the trustee in these circumstances. Such wishes may be overly onerous. I have suggested an upper bound for wishes with which a trustworthy person must comply. I have then further specified how to best interpret the wishes of the trustor in light of potential discrepancies between what he would have actually wanted, if consulted, and what he would have wanted had he deliberated in an ideal way. In the last section of the paper, I have argued that the existence of clear principles of action does not undermine the idea that such action can be discretionary as long as epistemic discretion is in place.

I conclude that all existing views of trustworthiness should be supplemented with TWA-UB.<sup>66</sup>

If what I say here is correct, then certain natural implications are also entailed for views of trust. While the exact nature of these implications is beyond the scope of the current paper, something along the following lines will be true: systematic failure to comply with TWA-UB in discretionary circumstances should (and most often will) undermine trust.

65. Thanks to Daniel Schwartz for pushing me to explore the connection between trustworthiness and friendship. No doubt much more than I have said here can and should be said about this connection.

66. As mentioned above, I suspect that most existing views are compatible with this supplement, though I am not committed to this claim.

## NOTES ON THE CONTRIBUTOR

Yonatan Shemmer is a senior lecturer at the University of Sheffield. His main interests are in metaethics and the philosophy of action. Some recent publications include *A Normative Theory of Disagreement* (2017), *Subjectivism about Future Reasons* (2019), *Disagreement without Belief* (2021), and *Disagreement for Dialetheists* (2024).