Philosophers' Imprint 17 (4):1-22 (2017)

Ryan Preston-Roedder
Occidental College

Civic Trust*

1

1.1 It is a commonplace that there are limits to the ways we can permissibly treat people, even in the service of good ends: we may not steal someone's wallet, even if we plan to donate the contents to famine relief; break a promise to help a colleague move, even if we encounter someone else along the way whose need is somewhat more urgent; or harvest organs from one person to save others who need transplants. In other words, we should observe *constraints* against mistreating people in certain ways, where a constraint is a moral principle that we should not violate, even when that is the only way to prevent further, similar violations or other, greater evils. But, despite its intuitive appeal, the view that there are constraints has drawn considerable criticism, and attempts to provide a rationale for constraints have been, at best, substantially incomplete.[1]

Discussions in the literature largely neglect a consideration that, I argue, is vital for fully understanding the justification and reason-giving force of constraints: whether someone is trustworthy depends largely on whether she observes constraints against mistreating people in certain ways.[2] Once we recognize the link between constraints and trust, we can articulate an important non-instrumental rationale for constraints. Roughly, observing con-

* The final publication is available at
http://hdl.handle.net/2027/spo.3521354.0017.004

[1] For influential criticisms of constraints, see Kagan 1989: Chs. 1, 3, and 4 and Scheffler 1994: Ch. 4. Below, I will discuss some of the main defenses of constraints in the literature.

[2] One discussion of constraints that emphasizes trust is Annette Baier's reading of Hume, according to which we should observe certain constraints, for example, a constraint against theft, because this is a means of producing a "climate of trust", and thereby promoting our self-interest (Baier 1994a: 11). My discussion is deeply indebted to Baier. But, unlike Baier, I argue that constraints have *non*-instrumental importance; furthermore, my account is as much inspired by Kant as by Hume.

straints is a condition for being worthy of a certain form of trust, which I call *civic trust*, and being worthy of such trust is an essential part of living with others in a form of harmony that characterizes morally permissible interaction.

Moral philosophers of more or less all sorts have accepted some version of the view that when someone observes moral principles, she thereby lives in a kind of harmony with others; furthermore, the view that appealing to some ideal of interpersonal relations is a promising approach to making sense of constraints has been "in the air for a while" (Kumar 2012). But the rationale I defend differs from accounts in the literature because it recognizes that living with people in the relevant sort of harmony involves trusting them in certain ways, and, crucially, adopting actions and attitudes that make it appropriate for them to trust *us* in certain respects.[3] This approach focuses on the role that observing moral principles plays in our psychological lives, and in the psychological lives of those around us.

This emphasis on the inner life is important, I argue, because it is vital for making sense of certain nuanced features of constraints, and accounting for constraints' reason-giving force. Furthermore, the approach I defend deepens our understanding of what it means to live with others in the kind of harmony that characterizes permissible interaction – what it means, in other words, to make something like the Kingdom of Ends a concrete reality. So attending to the link between constraints and trust yields advantages on two levels: first, it helps us make better sense of the justification and reason-giving force of certain widely held first-order moral judgments, and, second, it deepens our understanding of the familiar, plausible conception of morality to which I appeal to make sense of those first-order judgments.

1.2 I can more clearly describe these advantages if I first describe the main aim of a rationale for constraints, namely, to explain why our obligation to observe principles that have the distinctive structure of constraints makes sense in light of plausible background

[3] I argue below that living in such harmony with others involves behaving in ways that make it appropriate for them to trust us. In Preston-Roedder 2013: 683-685, I argue that living in such harmony with others involves being disposed to trust *them* in certain ways.

claims, for example, claims about the nature and point of morality. Philosophers often describe the main structural feature of constraints by appealing to the distinction between *agent-neutral* and *agent-relative* principles.[4] An agent-neutral principle directs everyone to whom it applies to promote the same aim, namely, producing some good occurrence or preventing a bad one; for example, a neutral principle that condemns coercion treats instances of coercion as bad occurrences and directs each of us to minimize such occurrences, even when this involves coercing one person in order to prevent others from being coerced.

By contrast, a constraint against coercion does not treat instances of coercion as bad occurrences to be minimized; rather, it prohibits each of us from coercing people, even when coercing someone is the only way to minimize instances of coercion. Constraints are agent-relative, and a relative principle assigns a different aim to each person to whom it applies; for example, a prudential principle directs each person to promote her own welfare, a parental obligation directs each to care for her own children, and a constraint against coercion prohibits each from coercing people herself.

The main aim of a rationale for constraints is to make constraints' relativity intelligible. But, to be clear, the fact that some principle is agent-relative is not, by itself, problematic; many principles that are central to our practical deliberation are agent-relative. Rather, constraints have additional features that make their relativity puzzling. The relativity of constraints against harm does not seem to derive from the importance, for potential victims, of avoiding harm. After all, someone who observes constraints will refrain from doing harm, even when harming someone is the only way to prevent more people from being harmed. On balance, she could better serve potential victims' interests by doing whatever would minimize the harm people suffer, and so concern for potential victims' interests may seem to support a neutral reason to minimize harm, rather than a constraint.

Nor does constraints' relativity seem to derive from more fundamental agent-relative reasons to protect our own interests, or the interests of people we care about. After all, a constraint prohibits us from doing harm whether or not observing this prohibition

[4] I use Derek Parfit's (1984: 27) characterization of this distinction.

would serve our own interests, and no matter what we care about. But if the relativity of constraints does not derive from the importance of protecting potential victims' interests, our own interests, or the interests of people we care about, then it may be unclear what its source could be. The main aim of a rationale is to explain how there could be prohibitions against harm that both prohibit us from doing what would best serve others' interests *and* bind us without regard to our particular aims.

One thing we may hope to accomplish by articulating such a rationale is convincing those who are skeptical or agnostic about our obligation to observe constraints that we should, in fact, observe them; and my account may serve this purpose to some degree. However, as I will explain below, some skeptics, including some Act Consequentialists, would likely reject one of the background claims on which my account relies, and so they may not find my account persuasive on its own. For that matter, I do not believe such skeptics would be persuaded by any of the main accounts in the literature.

But convincing skeptics that we should observe constraints is not my main motivation for articulating a rationale. Rather, I wish to make clearer, especially to agnostics and to those who already believe we should observe constraints, how constraints fit together with other practical principles, and with other values that have great moral significance. That is, I want to illuminate connections, which have been overlooked, between constraints and these other principles and values; and I wish to do this in a way that not only makes constraints' relativity intelligible, but also helps us better understand (1) the justification of other, more nuanced features of constraints (2) the reason-giving force of constraints and (3) the familiar conception of morality to which my account appeals.

1.3. There have been many attempts to make sense of constraints, and some of these attempts deepen our understanding of constraints, for example, by making their relativity more intelligible. But my account supplies part of the story that has been overlooked, and it thereby addresses important shortcomings of the main accounts in the literature.

One familiar approach to making sense of constraints is the Rule Consequentialist approach, which states that we should follow the principles adherence to which would produce the best results, and

that these principles include constraints.[5] I believe this approach supplies part of the rationale for constraints; nevertheless, our reason to observe constraints does not seem to depend primarily on the benefits we thereby produce. Suppose it turns out that our accepting a permission to imprison, without trial, people who are suspected of having committed violent crimes would produce somewhat better results, overall, than accepting a constraint against such imprisonment. Many judge that we should nevertheless observe the constraint; that is, the rationale for constraints seems to have an important non-instrumental component. This is the possibility I wish to explore.

There are also non-instrumental accounts of constraints in the literature. But even the most promising among them have important shortcomings: some lack resources needed to make sense of constraints' nuanced features, some rest on highly implausible assumptions, and some neglect aspects of moral life to which we must appeal to fully grasp constraints' reason-giving force. Below, I will explain how my account fills important gaps left by these non-instrumental approaches.[6]

My account makes constraints intelligible by identifying a morally significant relation that someone bears to people when, and only when, she observes certain constraints against mistreating them; and it characterizes this relation in a way that helps explain, in light of claims about the nature and point of morality, why it makes sense for a moral theory to include those constraints. The account rests partly on the view that it is sometimes reasonable for people to pursue their own private aims, rather than promote the greater good. For my purposes, claiming that someone's conduct is reasonable means that she has sufficient reason to adopt it, all things considered; and judgments concerning what it is reasonable take into account not only the agent's reasons to promote her own interests, but also reasons that derive from others' interests. It may be reasonable, in this sense, for someone to pursue a career she finds fulfilling, even if she could do more good by taking a

---

[5] See Hart 1961: 190-195 and Hooker 2000: Ch. 6.

[6] Recent non-instrumental accounts include Hurley 2009: Ch. 6; Kamm 1996: Ch. 10 and 2007: Chs. 1, 5, and 8; Kumar 1999: 304-309; Nagel 1995: 83-93; and Quinn 1993a.

more lucrative job and donating her surplus income to famine relief.

Though people can reasonably devote some special attention to their private aims, someone who observes moral requirements thereby limits her choice and pursuit of her own aims in ways that bring her life into a kind of harmony with other people's lives, provided that these others pursue reasonable aims by reasonable means. In other words, whatever further features they possess, moral requirements must be such that anyone who observes them thereby lives in substantial harmony with other people who pursue reasonable aims by reasonable means, even if those people's aims conflict, in limited respects, with hers. The view that interpersonal harmony characterizes morally permissible interaction is more or less universally accepted, in some form or other, and when it is properly understood, it forms the basis of the rationale I defend.

I argue that one condition someone must satisfy to live in such harmony with people is being worthy of their civic trust; roughly, she must behave in ways that make it appropriate for others to interact with her without fear, and to pursue their aims openly when they are around her. Provided that people can sometimes reasonably pursue their private aims, rather than promote the good, someone can be worthy of their civic trust only if she observes certain constraints against mistreating them. Since moral requirements must be such that anyone who observes them lives in harmony with others, it follows that these requirements must include constraints.

My task is to supply the main details needed to defend this rationale. In Section 2, I will discuss the importance of trust in social life, and explain more fully what it means to have civic trust in others and to be worthy of such trust oneself. In Section 3, I will discuss the connection between trustworthiness and constraints, arguing that – given some intuitively plausible assumptions – someone can be worthy of people's civic trust only if she observes certain constraints against mistreating them. In Section 4, I will discuss the basis of the rationale; that is, I will clarify and defend the view that when someone observes moral requirements, she lives in a kind of harmony with others, and I will explain more fully how to move from this virtual platitude to the judgment that we should observe constraints. Finally, in Section 5, I will briefly de-

velop my claim that each of the main accounts of constraints in the literature leaves out some important element of the rationale for constraints, an element that my account supplies.

2

2.1 Many of the projects and relationships that make life worth living flourish only in an atmosphere of trust. Trust is an essential part of the personal relationships that one finds in a good life. For example, the relationship between parents and children cannot thrive unless children trust their parents to support and protect them in certain ways, and parents trust their children, to some degree, to avoid certain dangers. Trust is also an essential part of good friendships. Friends tend to reveal private thoughts and feelings to each other, and each pursues ends that the friends share. So a friendship cannot thrive unless each friend trusts the other to refrain from belittling her ideas, undermining her private interests, or frustrating their shared ends.

Furthermore, to lead good lives, we must trust not only people who are close to us, but also people we do not know. Our flourishing depends partly on our capacity to trust the countless strangers we encounter every day; for example, riding the subway would be intolerable if we could not trust our fellow passengers, for the most part, not to steal our belongings or push us onto the tracks. Our flourishing also depends on our capacity to trust distant strangers whose behavior we could not monitor if we wanted to. We could not live in peace if we could not trust such strangers, for the most part, not to plant bombs on busy streets or poison our water supply.

Annette Baier (1994b: 95-110 and 1994c: 130-151) argues, plausibly, that the forms of trust that are most central to social life involve relying on people's goodwill toward us, rather than relying, say, on vigilance or threats, to ensure that those people do not harm us. For example, insofar as someone trusts his spouse to remain faithful, he will not interrogate her friends about her whereabouts, or threaten to abuse her if she has an affair. Rather, he will rely on her love for him to ensure that she remains faithful. And insofar as a parent trusts her teenager to make fairly responsible decisions, she will grant him some limited domain within which he can shape his own life, without interference. When we trust people in this way, we accept vulnerability to their power to harm us. So,

when the husband trusts his spouse, he accepts vulnerability to her power to make a fool of him, and to jeopardize a relationship that matters to him; and when the parent trusts her son, she accepts vulnerability to his power to cause her emotional anguish, and to squander opportunities that she struggled to provide for him.

2.2 These brief remarks leave us well placed to identify a form of trust, namely, civic trust, that someone makes appropriate when she observes certain constraints against harm; and they leave us well placed to appreciate respects in which our having such trust in one another, and being worthy of such trust ourselves, is morally significant. Roughly, someone who has civic trust in people is unafraid to interact with them, and she is open, or in other words, unguarded, in her dealings with them. She tends to pursue hobbies, cultivate relationships, engage in cultural or religious practices, and carry out her other projects – in short, she tends to live her life – without being wary or fearful. More precisely, someone who has civic trust in people is willing and unafraid (1) to interact with them, even if she is vulnerable to harm they might cause, and (2) to rely on their goodwill toward her – as opposed to relying on secrecy, force, or constant vigilance – to ensure that they do not use whatever they know about her in ways that harm her. Someone is worthy of civic trust just in case her character, which comprises entrenched dispositions of action, deliberation, and emotional response, does not, by itself, make it inappropriate for people to have such trust in her.

To be clear, when Baier claims that certain familiar forms of trust involve relying on people's goodwill toward us, she uses "goodwill" to refer to people's concern for us; by contrast, when I claim that civic trust involves relying on people's goodwill, I use that term more broadly, to refer either to people's concern or to their respect for us. To be *worthy* of people's civic trust, we must adopt prohibitions against mistreating them in certain ways, and limit our conduct and deliberation in accord with those prohibitions. We thereby grant people's interests considerable weight in our practical deliberation, exhibiting what Stephen Darwall calls "recognition respect" for them (Darwall 1977: 38).

Civic trust comprises two sets of attitudes and behaviors, which may come apart, and it will help to consider each separately. First,

someone who has civic trust in people is unafraid to interact with them, despite vulnerability to harm they might cause. She may be unafraid, say, to ride a city bus with them or stand in line with them in a crowded market. Such freedom from fear is appropriate just in case interacting with these people would not raise her total risk of being harmed, or suffering some other loss.

Second, someone who has such trust in people accepts vulnerability to their power to use what they know about her mind, body, or property in ways that make her worse off. She trusts people not only to refrain from raising her total risk of being harmed, but also to refrain from using what they know about her in ways that harm her. Such knowledge includes knowledge of her private thoughts and commitments, which others can acquire only by spending time with her, or with people who know her well. Beyond this, it includes knowledge of less personal facts, like the fact that she recently withdrew money from an ATM or that her drinking water comes from a certain reservoir, which others can acquire by observing her at the right moment, or by doing some research. And it includes knowledge of facts that may be derived from wholly general claims about human biology or psychology, like the fact that a sharp blow to the head may render her unconscious, or that she is unlikely to notice that a pickpocket is taking her wallet if his accomplice is distracting her.

This characterization implies that someone can be worthy of people's civic trust even if she is willing to harm them in certain limited ways. I can begin to explain which dispositions to harm make someone unworthy of civic trust, and which do not, if I compare two cases. First, imagine someone who steals from people when he wants something they possess and he is unlikely to get caught. Because he steals deliberately, for his own benefit, this person is unworthy of people's civic trust in both of two possible respects. That is, because he steals to promote his own interests, people raise their risk of suffering a loss when they interact with him, and so they have reason to avoid him, or fear that they will lose something valuable if they interact with him. Furthermore, this person is willing to use what he knows about people, for instance, the fact that someone's wallet is lying on a counter, or the fact that someone left her laptop in an unlocked office, to promote his own interests at their expense. So people have reason to prevent him from learning what valuable goods they possess and how these goods

may be reached, and they have reason to rely on vigilance or force to prevent him from using such knowledge when he acquires it.

By contrast, imagine that police are trying to capture a violent criminal who has fled, by car, from a crime scene. If the officers engage in a high-speed chase, they are likely to catch the criminal, but they might injure a bystander or damage a bystander's property during the pursuit. But if the officers do not engage in a high-speed chase, the criminal is likely to escape, and he may go on to commit additional violent crimes. Imagine that the officers are willing to engage in such a chase, despite the risk to bystanders, provided that this risk is "proportional" to the good they reasonably expect to accomplish.

The fact that the officers are so willing does not, by itself, make them less worthy of people's civic trust. Because the officers are willing to put bystanders at risk in this way only when this is necessary to secure greater benefits – in this case, capturing a violent criminal – this aspect of their characters does not, on balance, raise anyone's risk of being injured or suffering some other loss. So it does not give anyone reason, say, to try to prevent these officers from remaining part of the police force, or to worry if the officers get assigned to her district.

Furthermore, the officers are not disposed to use what they know about anyone's physical vulnerabilities, property, and so on in ways that harm her. For example, when they decide whether to engage in a high-speed chase, they treat the fact that the chase would expose some bystander to the risk of bodily injury or property damage solely as a reason *not* to pursue the criminal. Of course, this reason may not be decisive, whether it is considered on its own or together with their reasons to avoid exposing other bystanders to such risks. Nevertheless, it functions solely as a barrier to the pursuit. So no one has reason to avoid trusting these officers with facts concerning, say, her location or the value of her property, which the officers might use to help determine whether to pursue the criminal. More generally, the fact that the officers are willing to expose people to risks in this limited way does not, by itself, give anyone reason to prevent the officers from learning facts about her mind, body, or property; nor does it give anyone reason to prevent the officers from using such facts to carry out their aims. Below, I will describe in more detail ways in which

someone can be willing to harm people, while remaining worthy of their civic trust.

2.3 Civic trust and civic trustworthiness are both valuable relations. We have agent-neutral reasons to form communities whose members have such trust in one another. But we also have agent-*relative* reasons to be worthy of civic trust, whether or not this leads others to trust us; and some of these relative reasons are central to my rationale for constraints. The agent-neutral value of civic trust derives from the role that such trust plays in making life in a community worth living. People who have civic trust in one another are unafraid to live and work together, and a community thrives only if its members are, to a considerable degree, free from such fear. In Samuel Scheffler's words, fear "dominates" people, drawing their attention to the risk of injury or loss, and away from goods they might otherwise enjoy (Scheffler 2006: 4). Fear also isolates people, leading each to limit her contact with others or, in some cases, to withdraw from society almost entirely.[7] So fear not only prevents communities from thriving, but also, sometimes, prevents people from maintaining communities at all.

Someone who has civic trust in people also trusts them with knowledge of her activities, whereabouts, physical vulnerabilities, and so on. We cannot always prevent people from acquiring such knowledge, unless we withdraw from society altogether; and if we remain in society, the alternatives to such trust are grim. Someone who lacks civic trust in people around her might, as far as possible, pursue her aims in secret, working to prevent others from learning where she is, what she does, or what goods she possesses. She might use force to ensure that when people do learn something about her, they cannot use what they learn in ways that harm her. Or she might monitor people constantly to determine what they know about her and what they plan to do with this knowledge. In short, if members of a community lack civic trust in one another, they avoid living their lives openly, and each must endure a form of loneliness.

---

[7] Such considerations lead Hobbes to list "continual fear" among the "worst of all" the hardships that people endure in the state of nature (Hobbes 1651: Ch. 13, para. 9). Frederick Douglass, in an account of his escape from slavery, vividly describes the "great insecurity and loneliness" associated with such fear (Douglass 1994: 89 and 90).

We not only have agent-neutral reasons to promote civic trust, but also have agent-relative reasons to be worthy of such trust. Some of these relative reasons are instrumental, deriving from the importance, for each agent, of causing others to believe that she is trustworthy, and to treat her accordingly. But we also have relative reasons to be worthy of civic trust for its own sake, whether or not this leads others to trust us; and my argument rests on these *non*-instrumental reasons. We have these reasons because being worthy of civic trust is, by itself, an important part of living in a kind of harmony with others. In other words, if someone is *un*worthy of such trust, her character might give people reason to avoid her, or to worry that they will suffer a net loss when they share a park bench with her or stand in line with her at an ATM. Or it may give people reason to threaten her, hide from her, or monitor her to prevent her from using what she knows about them in ways that harm them. In short, she forms a gulf between herself and people around her, and being cut off from others in this way is, by itself, undesirable.

By contrast, if someone is worthy of civic trust, her character makes it appropriate for people in her community to ride a city bus with her or walk beside her in a public park, without worrying that they will suffer a net loss as a result. And her character makes it appropriate for these people to exhibit a kind of openness: they can appropriately live their lives without working constantly to prevent her from learning where they are or what they are doing, or from using such knowledge in ways that hurt them. In short, she lives in a kind of harmony with others. This harmony need not involve people's believing that she is trustworthy and treating her accordingly; rather, it is a *normative* relation that obtains just in case her character makes it appropriate for people to trust her in the ways I just described.

Living in such harmony with others is an essential part of interacting with them in a morally permissible way. It is morally desirable in itself, quite apart from whatever benefits it produces; indeed, the fact that someone who observes moral principles thereby lives in such harmony with others helps explain why observing such principles is worth caring about – why it is not just a matter of following pointless rules. In Section 4, I will develop the view that living in harmony with others is morally significant, and I will explain more clearly how this view provides the basis of a rationale

for certain constraints. But first, I will clarify the link between observing constraints and being trustworthy.

3

3.1 Our trustworthiness generally depends on the principles, or policies, that we observe. I argue that, to be worthy of people's civic trust, one must observe certain constraints against mistreating them. This is an argument about the kind of person someone must be in order to make it appropriate for others to feel and behave in certain ways, and so it relies on background claims about how people have reason to feel and behave. In particular, it relies on the intuitively plausible judgment that it is sometimes reasonable for people to pursue their private aims, rather than promote the greater good. To be clear, this is not the moral claim that people are sometimes morally permitted to pursue their private aims, but rather, the *non*-moral claim that they sometimes have sufficient reason to do so, all things considered.[8] It is possible, though rare, for someone to accept this latter claim, and nevertheless judge that people are always morally required to promote a single, overriding aim, say, the aim of securing the greatest happiness for the greatest number.[9]

To understand why my argument relies on this claim, consider two cases: First, suppose, contrary to what I believe, that people always have decisive reason to produce the best available results, impersonally considered. Now imagine that Jones does not observe any constraints; rather, he observes agent-neutral principles that always permit him to do whatever is needed to produce the best results, even when this involves harming one person to prevent others from being harmed.

It may be that, when someone recognizes that Jones is willing to harm her in order to help others, she will be afraid to interact with him. But, given that she always has decisive reason to promote the good, it is *in*appropriate for her to fear him, avoid him, or avoid trusting him with knowledge that he might use to harm her. In other words, if people always have decisive reason to do whatever is necessary to promote the good, then they have decisive reason

[8] For a recent defense of this claim, see Scheffler 2010.

[9] See Sidgwick 1907: Concluding Chapter.

to sacrifice their own interests whenever this will produce the best results. So Jones' willingness to harm people for the greater good does not, by itself, give them reason to avoid him, or to prevent him from using what he knows about them to carry out his aims.

By contrast, suppose it is sometimes reasonable for people to promote their private aims, and imagine, again, that Jones does not observe constraints against harm. In this second scenario, when someone reasonably pursues her own aims, it does not make sense, other things equal, for her to allow Jones to harm her, whether or not the harm is somehow necessary to produce the best results, impersonally considered. Put another way, there are cases where Jones' harming someone to promote the good would prevent that person from achieving some private aim that she reasonably pursues. In such cases, it would not make sense for Jones' potential victim simply to let Jones harm her. Rather, she might have reason to avoid Jones or to prevent him from learning where she is or how her property may be reached, and she has reason to be vigilant in order to prevent him from using such knowledge in ways that harm her. So the fact that Jones is willing to harm people in such cases makes him *un*worthy of their civic trust.

To be clear, Jones is untrustworthy, even if his willingness to harm one person in order to save others lowers everyone's risk of harm. Of course, if Jones' adherence to his principles lowers people's risk of harm, it does not give them reason to worry that they will suffer a net loss if they interact with him. Furthermore, if Jones becomes better able to help people as he learns more about their needs, then they have reason to let him learn about certain of their vulnerabilities, rather than living, as far as possible, in total secrecy. Nevertheless, anyone who does not wish to be sacrificed for the greater good also has reason to be discriminating in what she lets Jones learn about her. She has reason to prevent him from learning where she is, what she does, and so on when he is better able to use such knowledge to harm her than to help her; and even when she has reason to let Jones learn something about her, she has reason to rely on force or vigilance to ensure that he does not use this knowledge in ways that make her worse off. So, even when it makes sense for someone to *give* Jones information about her mind, body, or property, it does not make sense for her to *trust* him with it.

Given that people can reasonably promote certain private aims, one cannot be worthy of their civic trust if one is willing to do to them whatever is necessary to produce the best results. Rather, to be worthy of people's civic trust, one must observe certain agent-relative prohibitions against mistreating them. More precisely, one must observe three different prohibitions against harm, including two constraints; and, together, these prohibitions capture what I believe to be the central intuitions about the content of constraints against harm.[10]

First, to be worthy of civic trust, one must observe a prohibition against doing non-optimal harm, or in other words, harm that is not needed to produce the best results. This principle prohibits us from doing harm to others to promote our private aims, and from showing callous indifference to harm that our behavior may cause. To be clear, this principle is not, by itself, a constraint; to the contrary, it may be derived from the Act Consequentialist requirement that we always do what will produce the best results. But it is unsurprising that considerations that justify constraints – which sometimes prohibit harm that is necessary to promote the good – also justify a prohibition against non-optimal harm.

To show that observing this principle is a condition for being worthy of civic trust, I will consider someone who fails to observe it, and then determine how her willingness to violate the principle shapes her relation to others. Such a person may be willing to tarnish a colleague's reputation to advance her own career, or she may tend to hurt people's feelings or damage their belongings out of carelessness. Other things equal, her character gives people reason to restrain her or stay out of her way, and when people cannot avoid her, they have reason to worry that they will suffer a net loss as a result. So, when someone is willing to do harm that is not necessary to promote the good, this makes her unworthy of civic trust.

Second, to be worthy of civic trust, one must observe a constraint against intending harm, whether to promote one's own aims or to promote the greater good. To intend harm is to do harm or allow it to occur, either because the occurrence of harm is one's end or because it is a means of achieving one's end. When someone intends

[10] As I will explain below, these prohibitions make exceptions for certain kinds of harm.

harm, she aims to produce the harm, and this aim guides her conduct and deliberation. For example, someone who hurts a classmate's feelings to boost his own self-esteem intends the classmate's psychological distress as a means of promoting his own wellbeing. By contrast, consider a teacher who points out a mistake in her student's work in order to help him understand an assignment, even though this will embarrass him. She does not intend to cause him distress; rather, the student's distress is a foreseeable, but regrettable, side effect of her attempt to help.

Imagine someone who does not observe any constraint against intending harm – someone who will deliberately harm people, or allow them to be harmed, to achieve certain goals. To isolate the significance of her willingness to intend harm, as opposed to her willingness to do non-optimal harm, imagine that she intends harm only as a means of promoting the good. Even with this qualification, her willingness to intend harm makes her unworthy of civic trust. As she learns more about where people are, what goods they possess, and so on, she becomes better able to carry out her aim of harming them to promote the good; and so people who wish to avoid harm have reason to prevent her from acquiring such knowledge, or to use force or vigilance to prevent her from using it in ways that make them worse off.

To be clear, someone's trustworthiness may be undermined not only by her willingness to do harm deliberately, but also by her willingness to let harm occur, either because it is her end or because it is a means of achieving her end. Consider a doctor who is willing to let one patient die to provide organs for others who need transplants, and imagine that she learns that some patient, who has healthy organs, will die unless he receives treatment. The doctor will treat the fact that this patient has healthy organs, which can be used to aid others after he dies, as a justification for withholding treatment, and for avoiding conduct that would lead others to provide treatment. So it makes sense for patients to monitor her to determine how she uses such knowledge, and to use threats or coercion to prevent her from using such knowledge to justify letting them die.

Viewing the constraint against intending harm as a principle that someone must observe to be worthy of civic trust enables us to address an important difficulty in formulating that constraint: many

instances of intending harm can be described in such a way that, strictly speaking, the harm is merely foreseen, rather than intended.[11] Consider the familiar case where someone has life-threatening complications during labor. If the doctor performs a craniotomy, crushing the child's head, she will save the mother's life; otherwise, the child can be delivered unharmed, but the mother will die. Suppose the doctor performs the craniotomy. On the one hand, it seems that she intentionally kills the child to save the mother. On the other hand, one might claim that, strictly speaking, she does not intend to kill the child; rather, she intends to crush the child's head, and the death is merely a foreseen side effect. But, if we can deliberately crush someone's head without intending harm, then a constraint against intending harm is not nearly as morally significant as it initially appears to be.

One response to this difficulty is to insist that the doctor's aim of crushing the child's head is so closely related to the resulting death that, for purposes of moral evaluation, both should be considered part of her aim.[12] But, whether or not we accept this response, the constraint against intending harm that someone must observe to be worthy of civic trust avoids this problem. If someone is willing to crush people's skulls, whether to achieve her private aims or to promote the good, this gives people reason to fear her, avoid her, or prevent her from using what she knows about them to promote her aims. Her willingness to treat people this way makes her untrustworthy. We might say, adapting a formulation from Warren Quinn, that to be worthy of civic trust, someone must observe a constraint that not only prohibits her from intending harm, but also prohibits her from intentionally involving people in her pursuits in ways that result in their being harmed (Quinn 1993b).[13]

---

[11] See Bennett 1995: Ch. 11, Foot 2003: 21, McMahan 1994, and Quinn 1993b.

[12] See Foot 2003: 21-22.

[13] This does not mean that if the doctor performs the craniotomy, she thereby violates a constraint against intending harm. Constraints make exceptions for certain types of harm, and it may be that the constraint makes an exception for the craniotomy, say, because the child threatens the mother's life. My point is that deliberately crushing someone's head is the sort of act that this constraint *ordinarily* prohibits, whether or not one intends harm in some strict sense.

But, for ease of expression, I will describe this principle as a constraint against intending harm.

Third, to be worthy of civic trust, one must observe a constraint against doing harm that is a causal means, as opposed to a causal side effect, of promoting the good. This means that, on my account, both the distinction between intending and foreseeing harm *and* the distinction between doing and allowing harm are morally significant. As Shelly Kagan (1989: 87) notes, advocates of the one distinction often deny that the other distinction has any significance, but this seems misguided, because both distinctions are intuitively important. So the fact that my account provides a unified rationale for both distinctions is an advantage of the account.

Although intending harm and doing harm as a causal means are superficially similar, they are importantly distinct, and they can come apart. Someone who intends harm aims to produce harm, either as an end or as a means of achieving her end; but someone who does harm that serves as a causal means of producing some result need not aim at the harm or the result. The distinction between intending and foreseeing harm concerns an agent's aims, while the distinction between doing harm as a causal means and doing harm as a causal side effect concerns the role the harm plays in a causal chain leading from an agent to a result.

We can isolate the significance of someone's willingness to harm as a causal means if we consider someone who is willing to do such harm just in case that harm is *un*intended. Imagine that a nurse cares for a terminally ill patient who, despite his considerable pain, would value additional months of life. The nurse learns that a drug she administers to manage this patient's pain may hasten the patient's death, and that an alternative treatment would provide comparable pain relief without the associated risk. But, since the nurse is uncomfortable confronting the aggressive doctor who prescribed the drug, she decides not to pursue the matter, and to keep giving the potentially dangerous drug to her patient. Because some of the patient's vital organs have not yet been affected by his illness, they may be used, if he dies relatively quickly, to save others who need transplants. But making donor organs available is not among the nurse's aims; rather, she just wants to avoid an unpleasant confrontation with the doctor.

I can draw on my earlier discussion to show that the nurse's willingness to administer the drug, despite the risk to her patient, makes her unworthy of civic trust. Roughly, principles that permit the nurse to administer this drug are of two sorts, and observing principles of either sort makes one untrustworthy. On the one hand, the nurse's principle may permit her to do non-optimal harm. I explained above why observing such a principle makes someone untrustworthy. On the other hand, the principle might permit the nurse to harm people, provided that the harm would serve as a causal means of promoting the good. If the nurse observes this second type of principle, she treats the fact that some action, say, giving some drug to a patient, would cause harm as a defeasible reason not to perform it. So far, so good. But if that harm would somehow produce good results, she treats this further fact as a *justification* for doing harm. For example, when she learns that her patient has healthy organs, which may be used to save lives if he dies, she treats this as a justification for giving this patient a potentially lethal drug. So patients who have healthy organs, or other resources that may be used to help needy people if they die, have reason to prevent the nurse from learning that they have such resources, or to use vigilance or threats to prevent her from using such knowledge to justify conduct that would harm them.

3.2 Recognizing the link between constraints and trust, and appealing to this link to make sense of the structure of constraints, enables us to respond to one of the main challenges that accounts of constraints face. Typically, constraints prohibit us from doing or intending harm to others, but they make exceptions, the most important of which concern harm to people who reasonably consent to being harmed, harm in self-defense, and *un*intended harm that is a causal side effect, as opposed to a causal means, of promoting the good. So an intuitively plausible account of constraints must accommodate the view that constraints do not prohibit these types of harm. Applied to my account, this challenge takes the following form: my account rests on the view that, when someone acts rightly, she lives in a kind of harmony with others, where this involves being worthy of their civic trust; and so, to make room for exceptions to constraints against harm, I must show that someone can be worthy of people's civic trust, even if she is willing to do or intend relevant types of harm to them.

Turning first to consent, someone's willingness to harm people who reasonably consents to being harmed does not, by itself, make her untrustworthy. A clerk might sell cigarettes to someone who knows that smoking may cause cancer, but is willing to take the risk; a doctor might perform a preventive mastectomy on a patient who wishes to reduce her considerable risk of developing breast cancer; or someone might engage in limited forms of sexual sado-masochism with a willing partner. In such cases, someone's willingness to harm another person amounts to willingness to promote aims that, we can suppose, the other reasonably shares; and so her willingness to harm does not, by itself, give others reason to fear her, avoid her, or avoid trusting her with facts, say, about their bodies or property, that she might use to promote her aims.

It is more complicated to explain why someone's willingness to harm in self-defense need not make her untrustworthy. The fact that constraints make exceptions for harm in self-defense, including, perhaps, harm to morally blameless attackers, is among the least tractable features of constraints, and I do not have space to discuss this complicated topic here. So, although I believe this approach can help make sense of exceptions for harm in self-defense, I must set this issue aside.

Finally, consider the exception for doing unintended harm as a causal side effect of promoting the good. One widely discussed case to which this exception applies is the Trolley Case: a runaway trolley hurtles toward five workers who are repairing the tracks, and it will kill all five unless a bystander flips a switch that sends it to an alternate track, where it will kill one lone worker instead. Many judge that the bystander is permitted to turn the trolley, provided that her aim is to save the five, and killing the lone worker is merely a regrettable side effect. But, because the lone worker is an innocent person who wishes to live, defenders of constraints against doing harm, as opposed to intending it, have found it notoriously difficult to explain why the bystander is permitted to cause his death.[14]

Imagine that the bystander turns the trolley, and that causing the lone worker's death is merely a regrettable side effect, not part of

[14] For attempts to meet this challenge, see Foot 2003: 23; Kamm 1996: Chs. 6-7 and 2007: Ch. 5; Quinn 1993a; and Thomson 1976, 1985, and 1990: Ch. 7.

her aim. Provided that she observes the three prohibitions I described above, her willingness to harm people in cases like this, where the harm is an unintended result of promoting the good, does not make her untrustworthy. Because she observes a prohibition against doing non-optimal harm, she does not give transit workers any reason to worry, say, that they will suffer net losses if they repair the tracks at her local station. Because she observes a prohibition against intending harm, they need not worry that she will use what she learns about them to serve her aim of harming them. Finally, because she is willing to harm only as a causal side-effect, not as a causal means, of promoting the good, she does not treat the fact someone has a resource, which may be used to help others if he is harmed, as a justification for harming him. So the bystander's character does not give transit workers any reason to prevent her from learning where they are, what jobs they perform, or anything else she might use to determine whether her actions might harm them. Nor does her character give them reason to rely on force or threats to prevent her from using such knowledge to justify harmful behavior. To the contrary, each worker, including the one who ends up on the alternate track, has strong reasons to ensure that the bystander knows facts about his location, his physical vulnerabilities, and so on, which might help her determine whether her actions would harm him; and it is appropriate for each worker to trust the bystander with such knowledge.

Of course, the bystander may seem to treat one fact about the lone worker, not as a barrier to turning the trolley, but rather as a justification for doing so: the fact that he is alone on the alternate track. But the claim that some worker is alone runs together two considerations that, in this context, are importantly distinct. The first, namely, that the worker is on the track, concerns that very worker's location and vulnerability, but the second, namely, that no one else is on the track, concerns everyone else's location and vulnerability. Someone is worthy of civic trust when and because her character makes it appropriate for everyone around her to live *his own* life openly, without hiding facts about his own mind, body, or property from her, and without working to prevent her from using such facts to carry out her aims. This is the sense in which being worthy of someone's civic trust is part of living in harmony with *that person*. When the bystander decides whether to turn the trolley, and thereby cause some worker's death, she treats every relevant fact about that worker's mind, body, and

property as a reason not to turn it. So her willingness to turn the trolley does not give that worker, or anyone else, reason to avoid trusting her in relevant respects.[15]

I should add that, although the literature on this third exception focuses almost entirely on versions of the Trolley Case, there are other cases that have the same structure, and to which the exception applies. For example, officials might have to decide whether to allow a flood or wildfire to remain on its present course, or to direct it to another course, where it will harm fewer people. And Jonathan Glover describes a fascinating historical case with this structure: During World War II, British agents gained an opportunity to deceive the German military about the accuracy of their rocket attacks on London. Though the agents were not able to prevent the attacks altogether, they had the power to divert the rockets to less densely populated areas (Glover 1977: 102-103).[16] The same considerations that explain why someone who is worthy of

[15] Though my own intuitions about other, more complicated versions of the Trolley Case are not robust, many believe that the bystander is permitted to turn the trolley in Judith Thomson's Loop Case (Thomson 1985). In the Loop Case, as in the original Trolley Case, a runaway trolley will kill five people unless a bystander sends it to a side track; but, unlike the Trolley Case, the side track loops around to rejoin the main track. If there were no obstructions on the side track, there would be no reason to turn the trolley – it would simply loop around and hit the five from the other side. But there is one person on the side track, and if the trolley gets diverted, it will hit this person and stop. My approach may be unable to accommodate a permission to turn the trolley in the Loop Case, but I do not consider this a serious objection. Michael Otsuka points out that, unlike the original Trolley Case, "looping cases are not modeled after any real-word cases with which we are familiar and about which we have already formed reactions", and he appeals to some ingenious cases to explain away the intuition that turning the trolley is permissible in the Loop Case (Otsuka 2008: 109). Because my intuitions about the Loop Case and other, similarly complicated cases are not robust, and because the rationale I defend is intuitively plausible, I am content to let this rationale to guide my thinking about such cases.

[16] Warren Quinn (1993a: note 30) points out the similarity between this case and the Trolley Case.

civic trust may be willing to harm in the Trolley Case explain why she may be willing to harm in these other cases.[17]

3.3 In addition to making sense of these exceptions to constraints, my account can respond to another challenge that faces views, like mine, that aim to make constraints intelligible by appealing to the significance of interpersonal harmony. Such views must accommodate the judgment that we can permissibly engage in certain forms of competition, say, for jobs or awards, even though such competition seems initially to constitute a form of discord. In particular, my account, which grounds a rationale for constraints in the importance of being trustworthy, must accommodate the judgment that people engaged in certain forms of competition are permitted to act in ways that make them unworthy of competitors' trust in limited respects. Imagine that Smith lives in a community where many struggle to find work. Smith is permitted to apply for local jobs and present himself in a favorable light to potential employers. But his willingness to do this gives others reason to worry that he will take some job that they desperately need, and to avoid revealing facts about their own qualifications, which he might use to gain a competitive advantage.[18]

---

[17] An anonymous reviewer pointed out that, because our duty to observe constraints against harm is a correlate of people's *moral right* not to be harmed, this duty has another important structural feature: it is *directed* – we owe it to people to observe these constraints. I do not have space to discuss this directionality in detail, but my account helps make sense of it to some extent. One aspect of this directionality concerns the fact that people have limited powers to waive our duty to observe constraints against harming them. My account helps make sense of this power, because, as I explained above, it helps account for constraints' exceptions for harm to people who consent to being harmed. But another aspect of this directionality concerns the fact that people have authority to demand that we observe constraints against harming them. Such authority seems largely independent of the considerations I discuss; and so, to account for this authority, we must appeal to other claims about the nature and point of morality. For an overview of approaches to making sense of directionality, see May 2015. For an account of the authority to demand compliance with moral principles, see Darwall 2006. For discussion of the role that social practices play in conferring such authority, see Darby 2009.

[18] We seem permitted be untrustworthy in limited respects, not only when we compete for scarce resources, like jobs, but also, for example,

Generally, being worthy of people's civic trust is a condition for living with them in a form of harmony that characterizes permissible interaction. But this claim has an important caveat: in certain forms of competition, people can interact harmoniously, despite being untrustworthy in limited respects, because their outward conflict manifests some more fundamental unity. In other words, harmony and trustworthiness can come apart, to some degree, in such cases, and so people engaged in relevant forms of competition are permitted to act in ways that make them untrustworthy in limited respects.

I can explain how this applies to permissible competition for scarce goods if I briefly describe the conditions under which such competition often occurs. People pursue an enormous variety of projects. But there are not enough resources available for everyone to achieve all of her reasonable aims, and, morally speaking, no member of a society has a greater claim on that society's resources than any other member, other things equal. So everyone has decisive reason to adopt some fair procedure for distributing her society's resources, even though – crucially – this may sometimes result in other people's acquiring goods that she wants or needs.

Competition, together with limited forms of untrustworthiness that may accompany it, can be *un*objectionable on my account if it is part of some fair, efficient procedure for distributing a society's scarce resources. Provided that competition for jobs in Smith's community is fair and efficient, Smith's applying for jobs is, among other things, a reasonable way to promote an aim that he and other applicants have decisive reason to share: distributing their society's resources fairly. And promoting an aim that others share, or have decisive reason to share, is one way of being in harmony with them. To be sure, Smith's willingness to apply for jobs gives other applicants reason to distrust him in limited respects, but the conflict between Smith and his rivals rests on an

---

when we play competitive games. Since I do not have space to discuss all forms of permissible competition here, I will focus on competition for scarce resources, which seems less tractable than other forms.

underlying harmony grounded in their sharing, or having decisive reason to share, a single aim.[19]

To be clear, these considerations make room for permissible competition, but they do not justify permissions to violate the constraints I described above. I have argued that a person is permitted to act in ways that make her untrustworthy in limited respects, provided that her behavior promotes aims that all affected parties share, or have decisive reason to share. But when someone deliberately harms another person without consent, even if the harm is necessary to promote the good, he does not thereby promote any aim that he and his victim share, or should share, when the harm occurs. Of course, both agent and victim have reason to promote the good, but this reason is not decisive when the harm occurs – it is certainly not decisive *for the victim*. Rather, the victim has sufficient reason to promote her private interests, and when the agent tries to sacrifice those interests for the good, it is appropriate for her to resist, flee, and so on. So, unlike someone who is merely willing to compete for scarce goods, someone who is willing to harm people without consent, whether to promote his own aims or to promote the good, is untrustworthy in a sense that prevents him from living in harmony with others.

4

4.1 Now I can more clearly describe the rationale for constraints that I wish to defend. This rationale rests on the view that when someone observes moral requirements, she thereby lives in a kind of harmony with other people, provided that these people pursue reasonable aims by reasonable means. To be clear, this does not mean that someone who acts rightly lives in harmony with those who show callous disregard for others' interests; to the contrary, if someone, say, organizes a boycott to disrupt operations at a factory whose owners cruelly exploit their workers, then her actions may be both morally admirable and fundamentally discordant with the owners' aims. Rather, the view states that when someone observes moral requirements, she thereby lives in a kind of harmony with people who respond not only to reasons to promote

[19] Allan Wood (1999: 169-170) offers a similar account of how Kant's Kingdom of Ends formulation of the Categorical Imperative can accommodate permissible competition.

their own interests, but also to reasons to promote or protect others' interests.

This view about the relation between moral principles and interpersonal harmony is not part of an analysis of what we mean when we use terms like "moral principle"; rather, it is a moral judgement about one characteristic that the correct moral principles possess. Though I cannot provide a complete defense of this judgment here, I do present, in the course of my discussion, three considerations that support it. First, to make the intuitive appeal of this judgment more apparent, I will survey, in this section, some prominent moral theories, all of which adopt some version of the judgment, despite their substantial first-order disagreement about which behaviors are permissible, and their substantial metaethical disagreement about how our first-order judgments are justified. Second, also in this section, I will provide grounds for the judgment by appealing to claims about the character of people's interdependence, and about the significance of the view that each person is, in some sense, just one among others. Third, my particular interpretation of this judgment derives support from the role it plays in achieving reflective equilibrium among considered judgments about the cases that I discussed in Section 2 and 3, the structural features of constraints that I discussed in Sections 1 and 3, and the role that moral principles play in human life, which I will discuss below.

4.2 The view that when someone observes moral principles, she thereby lives in a kind of harmony with others is accepted, in some form or other, by a broad range of moral philosophers. Among the most influential statements of this view is Kant's *Kingdom of Ends* formulation of morality's ultimate principle, which states that when someone acts morally, she observes principles that she would endorse if she were a member of "a merely possible kingdom of ends" (Kant 1996a: 88). Kant describes this kingdom as "a systematic union of different rational beings through common laws", a union "of rational beings as ends in themselves and of the ends of his own that each may set for himself" (Kant 1996a: 83). So, on Kant's view, when people pursue their private aims within limits described by moral principles, their diverse aims are mutually supporting.

Contractualists also accept versions of this view. Contractualism states that someone acts rightly when and because she acts in accord with principles that any reasonable person would agree to adopt. On Thomas Nagel's characterization of Contractualism, part of what makes certain moral principles correct is the fact that observing them would realize a form of "harmony among the aims and actions of distinct persons" (Nagel 1991: 46). And in his Contractualist account of the reason-giving force of moral considerations, T.M. Scanlon claims that charges of immorality are distressing largely because of "their implication[s] for our relations with others, our sense of justifiability to or estrangement from them" (Scanlon 1998: 163). In other words, part of what is painful about someone's recognition that she has acted wrongly is her sense that she has disrupted her harmonious relation to others.

Mill's account of moral motivation supplies another, importantly different, formulation of this view (Mill 1863: Ch. 3). Mill claims that acts are right insofar as they promote happiness, where no one's happiness counts more or less than anyone else's; and he states that a natural source of motivation to act rightly, on this understanding of right action, is "the desire to be in unity with our fellow creatures" (Mill 1863: Ch. 3, para. 10).

This view that acting rightly involves living in some form of harmony with others makes sense, first of all, in light of the interdependence of people's lives. Of course, our lives are separate in one sense, because we often pursue our own private aims; but they are also interdependent in another important sense, because each person is vulnerable to the influence of countless others, and no one can survive, much less flourish, alone. As children, we cannot survive without the support of parents or the goodwill of strangers, and as adults, we pursue complicated, demanding aims, often lacking the resources needed to achieve our aims without help. And even if someone manages to survive and achieve her main aims more or less on her own, she is unlikely to live a *good* life unless she shares that life with people she cares about. The view that when a good person observes moral requirements, she thereby lives in a kind of harmony with others is fitting, given that people depend on one another in these ways.

This view also makes sense, given that moral principles must accommodate the judgment that everyone is, in some sense, just one

person among others. Put another way, there are a staggering number of people on the planet, and their various interests and aims come into sharp and frequent conflict. Each person's interests loom large from her own point of view, but, considered impersonally, no one's interests matter more or less than anyone else's, other things equal. Sensitivity to the fact that everyone is, in this sense, equal is an essential feature of moral principles, and moral principles accommodate this equality largely by requiring each person to behave in ways that bring her life into some form of harmony with others' lives. In other words, moral principles direct each person to limit her pursuit of her own aims in ways that bring her life into substantial harmony with other reasonable people's lives. And limiting each person's behavior this way – as opposed, say, to permitting her to pursue her own interests or the interests of people she cares about, without regard for others who may be affected – seems necessary for accommodating the judgment that each person is, in some sense, just one among others.

4.3 The view that we live in a kind of harmony with others when we observe moral requirements helps provide a rationale for constraints and helps account for constraints' reason-giving force, but only if we characterize this harmony in the right way. On one natural interpretation, which I reject, someone who acts morally lives in harmony with others by promoting the very same aim that everyone else promotes, or would promote if she were reasonable. Of course, there is no single, overriding aim that everyone actually promotes all the time. So this characterization of harmony is irrelevant, unless we assume that, all things considered, everyone always has decisive reason to promote a single, overriding aim, like the aim of producing the greatest good for the greatest number. But this assumption is intuitively implausible, and we need not accept it without argument. Instead, I accept the view that people sometimes have sufficient reason to pursue their private aims: reason to marry people they love, pursue careers they are passionate about, or take up hobbies they enjoy, even if they could do more good for more people by doing something else instead. On this alternative view, living in harmony with others cannot consist in promoting the same aim that everyone else promotes, or has decisive reason to promote, at all times. There *is* no such aim. Rather, it involves adopting and pursuing one's own aims in ways that, in some sense, leave room for others to pursue *their* reasonable aims, even when their aims differ from one's own.

I argue that someone can live in this latter sort of harmony with others only if she is worthy of their civic trust. To be clear, it may be that, in addition to being trustworthy, she must also satisfy other conditions, for example, observing principles that no one could reasonably reject, or being disposed to make personal sacrifices to meet others' basic needs. I find these additional suggestions plausible, but, as I said above, I cannot provide a complete characterization of such harmony here. Rather, I argue that, whatever else someone must do to live in such harmony with others, she must be worthy of their civic trust.

More precisely, when we consider what someone's life in a community is like when she is worthy of civic trust, and what it is like when she is not, we can recognize that being worthy of such trust is a condition for living in the relevant sort of harmony with others. Someone who is worthy of such trust acts and deliberates in ways that make it appropriate for others to take a walk in her neighborhood, sit beside her on a city bus, or ask her for directions, without fear. And she adopts and pursues her aims in ways that make it appropriate for others to pursue *their* reasonable aims openly, without guarding constantly against her learning where they are or what goods they possess, or guarding constantly against her using such knowledge in ways that hurt them. In short, it makes sense for people around her to adopt the sorts of attitudes and behaviors that not only make life in a community possible, but also make it worthwhile.

By contrast, someone who is *un*worthy of such trust behaves in ways that make it appropriate for people to avoid her, or to worry that they will be harmed if they happen to pass her on the street or stand beside her in a crowded market. Or she may behave in ways that make it appropriate for them to pursue their aims in secret, or to rely on threats or vigilance to prevent her from using what she learns about them in ways that harm them. Such a person is estranged from people around her, left to endure a profound form of isolation.

Put another way, the sort of harmony that characterizes morally permissible interaction does not just consist in our managing to avoid, for the most part, getting in one another's way; nor does it consist solely in our promoting one another's wellbeing. Rather, it also essentially concerns the quality of our psychological lives. To

live in the relevant sort of harmony with other people, someone must behave in ways that do not make it *in*appropriate for those people to have certain forms of trust in her – in ways that do not make it inappropriate for them to be part of her community, or to pursue their various projects in an open, unguarded way. Life in any physical community, for example, a neighborhood or workplace, that lacks this aspect of interpersonal harmony is obviously deeply impoverished. My point is that any conception of morally permissible interaction – that is, any conception of the *moral* community to which we should aspire – that overlooks this aspect of harmony is impoverished as well.[20] It overlooks considerations that are vital for fully grasping what it means to live in morally significant harmony with others, and for fully appreciating the appeal of such harmony.

4.4 The view that one must be worthy of people's civic trust in order to live in the relevant sort of harmony with them is, first of all, pivotal to my rationale for constraints. When we interpret harmony in this way, we supply a link between the virtual platitude that anyone who observes moral requirements thereby lives in a kind of harmony with others, and the judgment that moral requirements include constraints. I argued above that, given that people sometimes reasonably pursue their private aims, rather than promote the good, we cannot be worthy of their civic trust if we are willing to do to them whatever is necessary to produce the best results. Rather, to be worthy of their civic trust, we must observe constraints against doing and intending harm to them, constraints that have the features I described in Sections 1 and 3. So, provided that moral principles must be such that anyone who observes them lives in substantial harmony with others, and provided that being worthy of people's civic trust is a condition for living in such harmony with them, it follows that moral principles must include constraints with the features I described.

Furthermore, the link between observing constraints, being trustworthy, and living in harmony with others helps account for constraints' reason-giving force. As I said above, we can adopt two kinds of perspectives from which we can recognize what matters, and how we have reason to live. From the first, more personal perspective, our private interests loom large, and we can better appre-

[20] I am indebted to Geoffrey Sayre-McCord for this formulation.

ciate our reasons to devote some special attention to those private interests. But from the second, more detached perspective, we can better appreciate that there is some sense, central to the living of our lives, in which each of us is just one person among others – some sense in which *everyone's* life is enormously important, and no one's life is any more or less important than anyone else's.[21]

Given this background, the link between constraints, trust, and harmony helps account, in the following ways, for our reason to observe constraints. First, we have reason to guide and limit our conduct in ways that bring our lives into some form of harmony with other people's lives – as opposed, say, to showing callous indifference to others' lives – because this is part of responding appropriately to a form of value that everyone's life possesses; it is, in other words, part of recognizing and taking into account the fact that each of us is just one person among others. Second, we have reason to live with others in the particular form of harmony to which my account appeals – that is, reason to devote some special attention to our private aims, but do so in ways that leave room for others to pursue their reasonable aims – because this reconciles, to some degree, the two perspectives, which I just described, from which we determine how to live. By living in this form of harmony with others, we not only respond appropriately to a kind of value that everyone's life possesses in equal measure, but also accommodate the fact that our private interests loom large from our own points of view.[22] Third, part of living with people in this particular form of harmony is being worthy of their civic trust, or more precisely, having a character that makes it appropriate for them to live their lives openly, without being wary or fearful. When we are worthy of such trust, we thereby enter, in a limited but important respect, into a kind of community with others; but when we are *un*worthy of such trust, we are left to endure a profound form of isolation. The appeal of entering into this form of community, and avoiding the relevant form of isolation, helps account for the force of our reason to observe the principles – including the constraints – adherence to which makes us worthy of civic trust.

---

[21] See Nagel 1986: Ch. 9 and 1991: Chs. 1-5 and Scheffler 1994: Ch. 3.

[22] See Nagel 1991: Chs. 1-5.

4.5 We can most clearly articulate this account of the justification and reason-giving force of constraints when we assume that people are deeply invested in their private interests, and so, strongly averse to being harmed for someone else's sake. Now I want to consider, briefly, complications that arise when people's attachment to their private interests is somewhat attenuated.[23] Imagine that everyone reasonably devotes some special attention to her private interests, but nevertheless accepts that, because these are just the interests of one person among others, they may be sacrificed, on occasion, for the greater good. My account provides a rationale for constraints in this case, but these people's attitudes may limit, in ways that seem plausible on reflection, both the range of behaviors that constraints prohibit and the reason-giving force of constraints.

There are at least two senses in which such people might accept that their interests may be sacrificed. First, they might *endorse* having their interests sacrificed in certain cases; for example, they might endorse having their property seized for the greater good, though they object to having their bodies injured for that purpose. Suppose, for argument's sake, that these attitudes are reasonable. In that case, these people may consent to having their property seized in some broad class of cases and thereby establish, in the manner I described above, a broad class of exceptions to constraints against maltreating them. But, to be clear, to be worthy of these people's civic trust, one *still* has to observe constraints against maltreating them in other ways, to which they do not consent.

Second, even if these people object to having their interests sacrificed, they may not *care* much whether the sacrifices occur. For example, it may be that they generally object to having limited portions of their land seized for the greater good; nevertheless, because of the strength of their concern for the general welfare and the character of their concern for their private property, they do not care much whether such land seizures occur. Suppose, again, that their attitudes are reasonable. In that case, it may be that the reason-giving force of constraints against seizing portions of these people's land for the greater good is weaker, in some way, than it would be if they cared a great deal about avoiding such sei-

[23] An anonymous reviewer raised this issue.

zures. Roughly, the strength of these people's aversion to having portions of their land seized may help determine the *degree* to which someone's willingness to seize their land would make trust inappropriate, and undermine harmony. So, depending on how these considerations combine with other factors to determine the reason-giving force of moral principles – a topic I cannot address in detail here – the fact that these people do not care much whether portions of their land are seized may diminish, in some way, the reason-giving force of constraints against such seizures. This implication seems plausible. Constraints serve partly to limit our pursuit of our own aims in ways that accommodate other reasonable people's pursuit of their aims, and so it seems plausible that constraints' reason-giving force depends, in some way, on the degree to which these people care about securing their private aims.

5

I will close by returning to a topic I raised in the introduction: the relation between my account of constraints and the accounts in the literature. As I said above, the main accounts in the literature have important shortcomings. My account, which emphasizes, in ways these other accounts do not, the role that observing constraints plays in our psychological lives, corrects for many of these shortcomings; and it does so in ways that help us better understand the justification and reason-giving force of constraints, and better understand the familiar conception of morality to which my account appeals. To be clear, I do not aim, in this section, to raise *new* criticisms of accounts in the literature, but rather, to survey important criticisms that help clarify my account's contribution to our understanding of constraints.

Some accounts of constraints, including Rule Consequentialist accounts, are instrumental – they state that we have reason to observe constraints because this is a means of producing desirable results. I believe this approach supplies part of the rationale for constraints, but I deny that our obligation to observe constraints depends mainly on the benefits we thereby produce. The difference between my approach and the instrumental one is clearest in cases where everyone's accepting a permission to maltreat people in certain ways would have somewhat better results overall than everyone's accepting constraints against such maltreatment. I have argued that we should observe those constraints adherence to

which is part of being worthy of civic trust. Because outcomes in which people are trustworthy are desirable, instrumental accounts can appeal, in their own way, to claims about the link between constraints and trust. But people's trustworthiness is just one among many factors that make outcomes better. Outcomes might also be made better, say, by people's happiness, or by their freedom to direct their own lives. So there are cases where people observe constraints, and are therefore worthy of civic trust, but they nevertheless fail to produce the best available results. In such cases, my approach has more plausible implications than the instrumental approach.

Returning to an earlier example, suppose it turns out that everyone's accepting a permission to detain without trial people who are suspected of having committed violent crimes would reduce the rate of violent crime, and thereby produce somewhat better results overall, than everyone's accepting a constraint against such detentions. In that case, instrumental accounts imply that such detentions are permitted. By contrast, because our willingness to detain people in this way would make us unworthy of civic trust, my account implies, plausibly, that these detentions are prohibited. On my view, observing constraints is not just a matter of following principles adherence to which would have impersonally desirable results, or desirable results *for* people who observe them. Rather – and this point is crucial – it partly constitutes living in light of the recognition that, although each of us can reasonably devote some special attention to her private interests, each is, in one important sense, just one person among others.

There are also non-instrumental accounts in the literature, but these accounts, like their instrumental counterparts, are, at best, substantially incomplete. The most influential of these accounts can be divided into three categories. First, some accounts, which focus on making sense of constraints' agent-relativity, are poorly suited to make sense of constraints' other, more nuanced features. For example, Warren Quinn, Frances Kamm, and Thomas Nagel each argue that the fact that we are required to observe constraints against mistreating one another is somehow explained by the fact that we all have a certain desirable moral status, namely, the sta-

tus of inviolability, just in case we are so required.[24] But even if this curious account succeeds, it seems only to justify the view that we should observe constraints of some sort of other; after all, our being subject to any one of several different sets of constraints would render us inviolable in some respect, and to some degree.[25] So, it seems, we must appeal to other considerations to determine whether constraints prohibit us from doing harm or intending it, and, especially, to make sense of constraints' exceptions for permissible harm. These are among the issues my account addresses.[26]

Kant's moral theory provides the basis of a second category of non-instrumental accounts.[27] But Kant's theory has important shortcomings, and any account that relies on Kant's distinctive understanding of morality and rationality inherits these shortcomings. Kant claims that, to see ourselves as rational agents, as opposed to slaves of external forces, we must view a certain formal principle, namely, the Categorical Imperative, as the fundamental principle of practical reasoning; and he argues that familiar moral principles may be derived from the Categorical Imperative. But, like many people, I believe that the formal considerations to which Kant appeals are, by themselves, too restricted to yield a plausible moral theory; in particular, they are too restricted to yield a plausible account of the content of constraints.[28] In contrast to this

---

[24] See Quinn 1993a; Kamm 1996: Ch. 10 and 2007: Chs. 1, 5, and 8; and Nagel 1995: 83-93.

[25] For criticisms of this account, see Kagan 1991 and McNaughton and Rawling 1998. In Preston-Roedder 2014, I discuss the form of argument on which this account relies.

[26] Paul Hurley (2009; Ch. 6) defends another account in this category. Hurley's account – which revises and extends Scheffler's (1994: Chs. 2-3) rationale for moral permissions to promote one's private aims – appeals to claims about the moral significance of the personal point of view in order to provide a rationale for constraints. If Hurley's account succeeds, it makes the agent-relativity of constraints more intelligible; but it does not even purport to make sense of constraints' more nuanced features.

[27] I focus on Kant's statement of his views in the *Groundwork of the Metaphysics of Morals*.

[28] For recent discussion of this criticism, see Scanlon 2011: 117-126.

Kantian approach, my account rests on *substantive* moral judgments about the relation someone bears to people when, and only when, she acts rightly.[29]

Finally, Scanlon's Contractualism provides the basis of a third type of non-instrumental account (Scanlon 1998: Chs. 4-5). Contractualism's central claim, which can be developed in many different ways, is that we act rightly when and because we act in accord with principles that may be justified to others on grounds they cannot reasonably reject. Contractualist accounts of constraints rest on the view that constraints may be justified on such grounds. Contractualism, as Scanlon actually develops it, offers a striking contrast to Consequentialist thinking about right and wrong, largely because it stipulates two restrictions on the grounds that can justify moral principles: roughly, moral principles cannot be justified by claims about the impersonal value of outcomes, and they cannot be justified by claims about the combined weight of different people's reasons for preferring one principle to another (Scanlon 1998: 218-223 and 229-241). Some argue that, largely in virtue of these restrictions, Scanlon's view can provide a rationale for constraints.[30] But I believe these restrictions are implausible – claims

---

[29] Arthur Ripstein (2009: Chs. 1-2) develops a Kantian account of legal, as opposed to moral, constraints that appeals to Kant's political philosophy, rather than his moral philosophy; and one might think that Kant's political philosophy can ground an account of moral constraints as well. I do not have space to discuss this approach in detail, but one of its limitations is that Kant's political philosophy excludes an entire category of considerations that are central to understanding moral constraints' reason-giving force. Kant's political philosophy rests on his characterization of the relation we bear to one another when we observe just laws. For Kant, this relation consists solely in our limiting our outward conduct in ways that prevent us from getting in one another's way in certain respects (Kant 1996b: 23-24). But, as I argued above, an adequate characterization of the relation we bear to one another when we observe *moral* constraints focuses not only on our outward conduct, but also on our inner, psychological lives. We cannot fully grasp the sense in which observing moral constraints partly constitutes living in harmony with others – and so we cannot fully understand why observing those constraints is worth caring about – until we recognize, as my account does, that observing those constraints makes us worthy of certain forms of trust.

[30] See Kumar 1999: 304-309. Stephen Darwall claims that "it is well known that … constraints can be derived within Contractualism", and he

about the impersonal value of outcomes, and about the number of people who may be helped or harmed by our behavior, are sometimes directly relevant to the rightness or wrongness of actions.[31]

Unlike Scanlon's view, and unlike accounts of constraints that derive from it, my account accommodates the view that claims about impersonal value and claims about numbers are sometimes, by themselves, morally relevant. In other words, my account provides an alternative to Consequentialist thinking about constraints, but not, as Scanlon's Contractualism does, by simply ruling out as intrinsically irrelevant two of the main factors to which Consequentialist reasoning appeals. Rather, by appealing to judgments about the relation we bear to people just in case we act rightly, my account justifies limits on the ways these two factors help determine how we should act.

In this brief survey of accounts of constraints in the literature, I have described, in outline, some important criticisms of the accounts, and I have explained, in outline, how my account addresses these criticisms. But, to be clear, I do not claim that all of the accounts in the literature should be rejected entirely. To the contrary, as I have said throughout, some of these accounts supply important parts of the rationale for constraints. Nor, for that matter, do I claim that all of the comprehensive moral theories from which these accounts derive should be rejected entirely. I have argued that our obligation to observe constraints makes sense in light of claims about morality's nature and point, together with a claim about the kinds of aims we can reasonably pursue. But I believe that my account must, at the end of the day, be integrated into some more comprehensive understanding of morality, that is, into a theory or coherent set of theories that makes sense of constraints *and* provides rationales for other moral principles. It may be that my account should be integrated, say, into a substantially revised version of Kant's theory, or into a version of Contractualism that develops that theory's main idea differently than Scanlon does. But I cannot address this complicated issue here. My aim has been, not to refute accounts of constraints in the literature or to

---

attempts to give this approach "a deeper philosophical rationale" (Darwall 2006: 37).

[31] Parfit (2011: Ch. 21) develops this objection.

provide an alternative to comprehensive theories in the literature, but rather, to illuminate a certain value – namely, the value of civic trust – that we must appreciate in order to fully understand the justification and reason-giving force of constraints.[32] I have argued that recognizing this value enables us to better understand constraints and to better understand what it means to live with others in the kind of harmony that, according to a familiar, deeply plausible view, characterizes morally permissible interaction.

Acknowledgements

I am grateful to Anne Barnhill, Bernard Boxill, Adam Cureton, Winston Chiong, Don Garrett, Sara Goering, Peter Graham, Alex Guerrero, Elizabeth Harman, Thomas E. Hill, Jr., Frances Kamm, Douglas MacLean, Thomas Nagel, Derek Parfit, Erica Preston-Roedder, Jim Pryor, William Ruddick, Geoffrey Sayre-McCord, Joshua Schechter, Seana Shiffrin, Sharon Street, Larry Temkin, Simon Timm, Peter Unger, J. David Velleman, and two anonymous readers for *Philosophers' Imprint* for enormously helpful comments. I also presented portions of this paper to audiences at several department colloquia, where I received many helpful comments. This project was made possible through the support of a grant from Templeton Religion Trust. The opinions expressed in this project are those of the author and do not necessarily reflect the views of Templeton Religion Trust.

References

Baier, Annette. 1994a. What Do Women Want in a Moral Theory?, in *Moral Prejudices*. Cambridge, MA: Harvard University Press, 1-17.

_____. 1994b. Trust and Antitrust, in *Moral Prejudices*. Cambridge, MA: Harvard University Press, 95-129.

_____. 1994c. Trust and Its Vulnerabilities, in *Moral Prejudices*. Cambridge, MA: Harvard University Press, 130-151.

Bennett, Jonathan. 1995. *The Act Itself*. Oxford: Clarendon Press.

Darby, Derrick. 2009. *Rights, Race, and Recognition*. Cambridge: Cambridge University Press.

Darwall, Stephen. 1977. Two Kinds of Respect. *Ethics* 88: 36-49.

_____. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press.

[32] I am indebted to Geoffrey Sayre-McCord for this formulation.

Douglass, Frederick. 1994. *Narrative of the Life of Frederick Douglass, an American Slave*, in *Frederick Douglass: Autobiographies*, ed. Henry Louis Gates, Jr. New York: Literary Classics of the United States, Inc.

Foot, Philippa. 2003. The Problem of Abortion and the Doctrine of the Double Effect, in *Virtues and Vices: And Other Essays in Moral Philosophy*. Oxford: Oxford University Press, 19-32.

Glover, Jonathan. 1977. *Causing Death and Saving Lives*. London: Penguin Books.

Hart, H.L.A. 1961. *The Concept of Law*. Oxford: Clarendon Press.

Hobbes, Thomas. 1651. *Leviathan*.

Hooker, Brad. 2000. *Ideal Code, Real World*. Oxford: Oxford University Press.

Hurley, Paul. 2009. *Beyond Consequentialism*. Oxford: Oxford University Press.

Kagan, Shelly. 1989. *The Limits of Morality*. Oxford: Clarendon Press.

_____. 1991. Response to My Critics. *Philosophy and Phenomenological Research* 51: 919-928.

Kamm, Frances. 1996. *Morality, Mortality: Volume II: Rights, Duties, and Status*. Oxford: Oxford University Press.

_____. 2007. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. Oxford: Oxford University Press.

Kant, Immanuel. 1996a. *Groundwork of the Metaphysics of Morals*, in *Practical Philosophy*, ed. and trans. Mary J. Gregor. Cambridge: Cambridge University Press, 37-108.

_____. 1996b. *The Metaphysics of Morals*, trans. Mary J. Gregor. Cambridge: Cambridge University Press.

Kumar, Rahul. 1999. Defending the Moral Moderate: Contractualism and Common Sense. *Philosophy and Public Affairs* 28: 275-309.

_____. 2012. Review of Paul Hurley, *Beyond Consequentialism*. *Notre Dame Philosophical Reviews*.

May, Simon Cabulea. 2015. Directed Duties. *Philosophy Compass* 10: 523-532.

McMahan, Jeff. 1994. Revising the Doctrine of Double Effect. *The Journal of Applied Philosophy* 11: 201-212.

McNaughton, David, and Piers Rawling. 1998. On Defending Deontology. *Ratio* 11: 37-54.

Mill, John Stuart. 1863. *Utilitarianism*.

Nagel, Thomas. 1986. *The View From Nowhere*. Oxford: Oxford University Press.

_____. 1991. *Equality and Partiality*. Oxford: Oxford University Press.

_____. 1995. Personal Rights and Public Space. *Philosophy and Public Affairs* 24: 83-107.

Otsuka, Michael. 2008. Double-Effect, Triple-Effect, and the Trolley Problem. *Utilitas* 20: 92-110.

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

_____. 2011. *On What Matters: Volume II*. Oxford: Oxford University Press.

Preston-Roedder, Ryan. 2013. Faith in Humanity. *Philosophy and Phenomenological Research* 87: 664-687.

_____. 2014. A Better World. *Philosophical Studies* 168: 629-644.

Quinn, Warren. 1993a. Actions, Intentions and Consequences: The Doctrine of Doing and Allowing, in *Morality and Action*. Cambridge: Cambridge University Press, 149-174.

_____. 1993b. Actions, Intentions and Consequences: The Doctrine of Double Effect, in *Morality and Action*. Cambridge: Cambridge University Press, 175-193.

Ripstein, Arthur. 2009. *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge, MA: Harvard University Press.

Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

_____. 2011. How I Am Not a Kantian, in Derek Parfit. *On What Matters: Volume II*. Oxford: Oxford University Press, 116-140.

Scheffler, Samuel. 1994. *The Rejection of Consequentialism*, Revised edition. Oxford: Clarendon Press.

_____. 1996. Is Terrorism Morally Distinctive? *The Journal of Political Philosophy* 14: 1-17.

_____. 2010. Morality and Reasonable Partiality, in *Equality & Tradition: Questions of Value in Moral and Political Philosophy*. Oxford: Oxford University Press, 41-75.

Sidgwick, Henry. *The Methods of Ethics*, Seventh edition. London: Macmillan and Company, Limited.

Thomson, Judith Jarvis. 1976. Killing, Letting Die, and the Trolley Problem. *The Monist* 59: 204-217.

_____. 1985. The Trolley Problem. *Yale Law Journal* 94: 1395-1415.

_____. 1990. *The Realm of Rights*. Cambridge, MA: Harvard University Press.

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