

Ryan Preston-Roedder  
UNC Chapel Hill

## **Civic Trust\***

### I

It is a commonplace that we should observe deontological constraints against treating people in certain ways, where a constraint is a moral principle one should not violate, even when this is the only way to prevent further, similar violations or other, greater evils. In short, there are limits to the ways we can permissibly treat people, even in the service of good ends: one may not steal someone's wallet, even if one plans to donate the contents to famine relief; break one's promise to help a colleague move, even if one encounters someone else whose need is somewhat more urgent; or harvest organs from one person to prevent the deaths of others who need transplants. 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 substantially incomplete at best.<sup>1</sup>

---

\* 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, and J. David Velleman for helpful comments and discussion. I presented portions of this paper to audiences at several department colloquia, where I received many helpful comments.

<sup>1</sup> For influential criticisms of constraints, see Kagan 1989: Chs. 1, 3, and 4 and Scheffler 1994: Ch. 4. Later in this section, and in Section V, I will discuss some of the main defenses of constraints in the literature.

Discussions in the literature largely neglect a consideration that is vital for fully understanding the justification and reason-giving force of anything like the constraints that Common-Sense Morality comprises: whether someone is trustworthy depends, in part, on whether she observes constraints against mistreating people in certain ways.<sup>2</sup> Once we recognize this link between constraints and trust, we can identify a certain non-instrumental rationale for constraints. This rationale states, roughly, that observing constraints is a condition for being worthy of a particular form of trust, which I call *civic trust*, and that being worthy of such trust is an essential part of living with others in the sort 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. And 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 part of living with people in the relevant sort of harmony consists in trusting them in certain ways, and, crucially, behaving in ways

---

<sup>2</sup> One related discussion that emphasizes trust is Annette Baier’s reading of Hume, according to which our obligation to observe certain deontological prohibitions, for example, prohibitions against theft and breaking promises, derives from the fact that people’s observing these prohibitions is a means of producing a “climate of trust”, and so, a means of promoting their self-interest (Baier 1994a: 11 ). My discussion is deeply indebted to Baier. But, unlike Baier, I argue that constraints have *non-*instrumental importance, and I develop an account that is as much inspired by Kant as by Hume.

that make it appropriate for them to trust her in certain respects.<sup>3</sup> This approach focuses on the role that observing moral principles plays in someone's psychological life, and in the psychological lives of those around her.

As I will explain, this emphasis on the inner life is important, first of all, because it enables us to make sense of certain nuanced features of constraints, and to account for constraints' reason-giving force, better than if we overlook this connection between constraints and trust. Furthermore, the approach I defend enriches our understanding of what it means to live with others in the kind of harmony that characterizes permissible interaction, or in other words, what it means to make something like a Kingdom of Ends a concrete reality. This means that attending to the link between observing constraints and being trustworthy yields advantages on two levels: First, it enables us to make better sense of the justification and reason-giving force of certain widely accepted judgments about how to act. Second, it deepens our understanding of a familiar conception of morality to which I appeal to make sense of those first-order judgments.

I can more clearly describe the merits of this approach if I first describe the main aim of a rationale for constraints, namely, to explain why our obligation to observe principles with the structure of constraints makes sense in light of intuitively plausible background claims, including claims about the nature and point of morality. In other words, the aim is to explain, by appealing to such background claims, why it

---

<sup>3</sup> Later in this section and in Section IV, I defend the view that part of living in this sort of harmony with people is behaving in ways that make it appropriate for them to trust *us*. In Preston-Roedder 2013, I defend the view that part of living in this sort of harmony with people is being disposed to trust *them* in certain ways.

is rational for a moral theory to include principles with this structure. Philosophers typically describe the main structural feature of constraints – constraints also have more nuanced features that I will discuss in Section III – by appealing to the distinction between agent-neutral and agent-relative principles.<sup>4</sup> An agent-neutral principle directs everyone to whom it applies to promote the same aim, namely, producing good occurrences or preventing bad ones. For example, an agent-neutral principle that condemns coercion treats instances of coercion as bad occurrences and directs each of us to minimize coercion, 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 others, even when coercing someone is the only way to minimize instances of coercion. Constraints are agent-relative, and an agent-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 requires each to care for her own children, and a constraint against coercion prohibits each from coercing people herself.

So, part of the aim of a rationale for constraints is to make the relativity of constraints intelligible. To be clear, the mere fact that some principle is agent-relative is not, by itself, problematic. Many of the reasons that are central to people's deliberation are agent-relative: reasons to protect one's children from harm, to help one's friends achieve their aims, and so on. But other features of constraints make

---

<sup>4</sup> I use Derek Parfit's (1984: 27) characterization of this distinction.

their relativity puzzling, at least initially. The relativity of constraints against harm does not derive in any obvious way from the importance, for potential victims, of not being harmed. After all, someone who observes constraints will not do harm, even when harming one person is the only way to prevent more people from being harmed. On balance, she could better serve potential victims' interests by doing what would minimize the total harm people suffer. So, concern for potential victims' interests may seem to support a neutral reason to minimize harm, rather than a constraint.

Nor does the relativity seem to derive from more fundamental agent-relative reasons to protect one's own interests, or interests of people one cares about. After all, the constraint prohibits one from doing deliberate harm whether or not observing this prohibition would serve one's own interests, and no matter what one cares about. But if the relativity of constraints does not derive from the importance of protecting victims' interests, one's own interests, or interests of people one cares about, it may seem unclear what its source could be. Part of the 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.

Of course, one thing philosophers might hope to achieve by providing such a rationale is to convince people who are skeptical or agnostic about our obligation to observe constraints that we should, in fact, observe them. The account I defend can serve this purpose, but only in a limited way. Some agnostics would accept the background claims on which my account relies, and so my account would likely convince them that we should observe constraints. By contrast, certain skeptics, for example, certain Act Consequentialists, would likely reject one of these claims, namely, my claim, which I will discuss below, about the kinds of aims people have

sufficient reason to pursue all things considered. So, my account may not persuade them on its own. For that matter, I do not believe any of the main accounts in the literature are likely to persuade such skeptics that we should observe constraints.

But convincing skeptics that we should observe constraints is not the only reason to provide a rationale, nor is it among my main reasons for doing so. Rather, I wish to make clearer – especially to agnostics and to people who believe we should observe constraints of some sort or other – how constraints fit together with other principles of practical reason, and with values that have deep moral significance. In other words, I want to illuminate connections, which have been overlooked or poorly understood, between constraints and these other principles and values. Furthermore, I want to do this in a way that not only makes constraints' agent-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 conception of morality to which my account appeals.

The literature includes many attempts to provide a rationale for constraints. I believe some of these accounts make substantial contributions to our understanding of constraints; in particular, some may help make the relativity of constraints more intelligible. But my account supplies an important part of the story that has been left out so far, addressing shortcomings of each of the main accounts in the literature.

The most familiar approach to making sense of constraints is the Rule Consequentialist approach, which states, first, that we should follow the set of principles adherence to which would produce the best results, and, second, that this

set of principles includes constraints.<sup>5</sup> I believe this approach supplies part of the rationale for constraints, but even so, our reason to observe constraints does not seem to depend entirely, or even primarily, on the benefits we would 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 our accepting a constraint against such imprisonment. Many people judge that we should nevertheless observe the constraint. In other words, the rationale for constraints seems, to many people, to have an important, non-instrumental component. This is the possibility I wish to explore.

Of course, there are many non-instrumental accounts in the literature, but even the most promising among them have important shortcomings: some lack the resources needed to make sense of constraints' nuanced features, some rely on highly implausible assumptions, and some neglect aspects of moral life to which we must appeal to fully grasp the reason-giving force of constraints. In Section V, after I have presented my own account, I will explain how my approach fills important gaps left by non-instrumental accounts in the literature.<sup>6</sup>

The approach I adopt makes constraints intelligible by identifying a morally significant relation someone bears to people when, and only when, she observes constraints against mistreating them in certain ways. That is, it describes the relation someone bears to people just in case she observes certain constraints, and it does so

---

<sup>5</sup> See Hart 1961: 190-195 and Hooker 2000: Ch. 6.

<sup>6</sup> 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.

in a way that helps explain, in light of claims about morality's nature and point, why it is rational for a theory to include such constraints. The version of this approach that I defend rests on the view that it is sometimes reasonable for people to pursue their own private aims, even though this would prevent them from producing the best available results impersonally considered. For my purposes, it is reasonable for someone to behave in some way just in case she has sufficient reason to behave in that way, all things considered. This judgment takes into account not only the person's reason to promote her own interests, but also reasons that derive from other people's interests. It may be reasonable in this sense for someone to marry whomever she loves and pursue whatever career interests her, even if she could do more good for more people by remaining single, devoting herself to some more lucrative career, and donating her surplus income to famine relief.

Nevertheless, when someone observes moral requirements, she 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 people 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 people around her who pursue reasonable aims by reasonable means, even if these other people's aims conflict in certain limited ways with her own. 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 will argue that one condition someone must satisfy in order to live in such harmony with people is being worthy of people's 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 reasonably pursue some of their private aims, someone can be worthy of people's 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 certain constraints.

Now I will supply the main details needed to defend this rationale. In Section II, I will briefly discuss the importance of trust in social life, and I will explain in more detail what it means to have civic trust in others, and to be worthy of such trust. In Section III, I will discuss the connection between trustworthiness and constraints, defending the claim that – given certain plausible assumptions – one can be worthy of people's civic trust only if one observes certain constraints against mistreating them. In Section IV, 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 harmony with others, and I will explain more fully how to move from this platitude to the judgment that we should observe constraints against harm. Finally, in Section V, I will briefly develop my claim that each of the main accounts of constraints in the literature leaves out some important element of the rationale for constraints, some element that my account supplies.

## II

Most of the projects and relationships that make human lives worth living flourish only in an atmosphere of trust. To begin with, trust is an essential part of the intimate relationships one expects to find in a good life. The relationship between parents and children cannot thrive if children do not trust their parents to provide

them with certain forms of support, or if parents do not trust their children, to some degree, to avoid certain dangers. Trust is also essential to good friendships. Good friends sometimes reveal private thoughts, feelings, and aims to each other, and each devotes energy to projects that the friends pursue together. So a friendship cannot thrive unless each friend trusts the other not to belittle her ideas, undermine her private aims, or frustrate their shared pursuits.

To lead good lives, we have to trust not only people who are close to us, but also people we do not know. Our flourishing depends partly on our ability to trust the countless strangers we encounter each day. Riding the subway to work would be intolerable if we could not trust other passengers, for the most part, not to steal our belongings or push us onto the tracks. Our flourishing also depends on our ability to trust distant strangers whose behavior we cannot monitor even if we want to. We cannot live our lives in peace if we are unable to trust distant strangers, for the most part, not to plant bombs on busy streets or poison our water supply. And even when we rely on the justice system to prevent such evils, we trust distant police officers, judges, and lawmakers to protect us, rather than exploiting or neglecting us.

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 be faithful, he will not rifle through her cell phone records, interrogate her friends about her whereabouts, or threaten to abuse her if she flirts with other men. Rather, he will rely on her love for him, and her desire to maintain the relationship, to ensure that she does not have an affair. And insofar as a parent trusts her teenage children to

make fairly responsible decisions, she will give them some limited domain within which they can begin to shape their own lives, 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 to be faithful, he accepts vulnerability to her power to make a fool of him, to cause him grave distress, and to jeopardize a relationship that matters to him. And when the parent trusts her children, she accepts vulnerability to their power to cause her emotional anguish, and to squander opportunities she has struggled to provide for them.

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. They also leave us well placed to appreciate respects in which our having such trust in one another, and being worthy of such trust ourselves, are morally significant. Describing what it means to be worthy of civic trust amounts to describing the relation someone bears to people just in case she observes certain prohibitions – including certain constraints – against mistreating them. With this description in place, we can better understand why our obligation to observe constraints makes sense in light of deeper claims about the nature and point of morality, and we can determine, with greater clarity and confidence, which behaviors these constraints rule out.

Roughly, someone who has civic trust in people is not afraid to interact with them or to be open – or in other words, unguarded – in her dealings with them. When she is part of a community of people in whom she has such trust, she is disposed to pursue hobbies, cultivate friendships, engage in cultural or religious practices, and carry out projects – in short, she is disposed to live her own 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 – rather than 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.

Civic trust comprises two sets of attitudes and behaviors, which can come apart, and it will help to consider each set separately. First, someone who has civic trust in people is not afraid to interact with them, despite her vulnerability to harm they might cause. She may be unafraid to ride a city bus with them, ask them for directions, or stand in line with them at a crowded market. Such freedom from fear is appropriate just in case her interacting with these people would not raise her total risk of being harmed, or suffering some other loss.

Second, someone who has civic 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 commitments and private thoughts, which others can acquire only by spending time with her, or with people who know her. But it also includes knowledge of less personal facts – like the fact that she has just withdrawn money from an ATM or the fact that her drinking water comes from a certain reservoir – which others can acquire simply by observing her at the right

moment or by doing a little research. And it even includes 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 the fact that she is less likely 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 seems unlikely to get caught. Because he steals deliberately, and does so 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. So they have reason to avoid him, or to fear that they will lose something valuable when 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 has 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 can 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 officers are trying to arrest some violent criminal who has fled, by car, from the scene of a crime. If the officers engage in a high speed chase, they are likely to catch the criminal, but they might injure a bystander or damage some 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 may commit more 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 can reasonably expect to accomplish, and provided that this risk is an unavoidable aspect of the chase.

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 to secure a greater benefit – namely, getting a violent criminal off the streets – 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 them from joining the police force, or to worry if these officers happen to be assigned to her district.

Furthermore, the officers are not disposed to use what they know about anyone’s patterns of behavior, physical vulnerabilities, property, or the like in ways that harm her. For example, when the officers must decide whether to engage in a high speed chase, they treat the fact that the chase would expose some bystander to the risk of being injured, or having some valuable piece of property damaged, solely as a reason not to pursue their suspect. Of course, this reason may not be decisive, whether it is considered on its own or together with the officers’ 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 the officers with facts concerning, say, her location or the value of her property, which the officers might use to help determine whether to pursue their suspect. 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 own mind, body, or property, nor does it give anyone reason to prevent the officers from using such facts to carry out their aims. In Section III, I will discuss in greater detail ways in which someone can be willing to harm people, while remaining worthy of their civic trust.

Civic trust and civic trustworthiness are both valuable relations. People have strong agent-neutral reasons to form communities whose members have such trust in one another. But people also have strong agent-*relative* reasons to be worthy of civic trust, whether or not this leads others to trust them, and some of these relative reasons are central to my rationale for constraints. The agent-neutral value of civic trust derives from the role such trust plays in making life in a community worth living. Someone who has civic trust in people around her is not afraid to live or work with them, and a community can thrive only if its members are free from such fear. In Scheffler's words, fear "dominates" people, drawing their attention to 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.<sup>7</sup> So, fear not only prevents communities from thriving, but also, in some cases, prevents people from forming or 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,

---

<sup>7</sup> These are among the factors that lead Hobbes to list "continual fear" among the "worst of all" the hardships people endure in the state of nature (Hobbes 1651: Ch. 13, para. 9).

and if we remain in society, the alternatives to such trust are grim. Someone who lacks civic trust in people around her might pursue her aims in secret, preventing 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, they cannot use what they learn in ways that harm her. Finally, she might monitor people constantly to determine what they know about her and what they plan to do with the knowledge they have. So, if members of a community lack civic trust in one another, they avoid living their lives openly, and each endures a form of loneliness.

People 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, and they derive from the importance, for each agent, of causing others to believe that she is trustworthy, and to treat her accordingly. But people also have relative reasons to be worthy of civic trust for its own sake, whether or not this causes others to trust them, and my argument rests on these *non-instrumental* reasons. We have these reasons because being worthy of civic trust is itself an important part of living in harmony with people. In other words, if someone is unworthy of such trust, her character might give others 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 front of her in line at an ATM. Or it might 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, in itself, undesirable.

By contrast, if someone is worthy of civic trust, her character makes it reasonable, other things equal, for people in her community to ride a city bus with

her or share a park bench, without worrying that they will suffer a net loss as a result. Her character also makes it appropriate for people in her community to exhibit a kind of openness. They can reasonably 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 harmony with others. This harmony need not involve people's believing that she is trustworthy or 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 ways I 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 this sort of harmony with others helps explain why observing these principles is worth caring about, why it is not just a matter of following pointless rules. In Section IV, I will develop the view that living in harmony with others is morally significant, and I will explain in more detail how this view provides the basis of a rationale for certain constraints. But before I turn to these tasks, I will clarify the link between observing constraints and being trustworthy.

### III

Whether someone is trustworthy depends largely on which principles guide her behavior. I argue that, to be worthy of people's civic trust, one has to observe constraints against mistreating them in certain ways. This is an argument about the kind of person one must be in order to make it appropriate for people to feel and behave in ways I described above. So, it relies on background claims about how people have reason to feel and behave. In particular, it relies on the intuitively

plausible claim that it is sometimes reasonable for people to pursue their own aims, rather than promote the good. To be clear, this is not the moral claim that people are sometimes morally permitted to pursue their own aims, but rather, the non-moral claim that people sometimes have sufficient reason to pursue their own aims, all things considered.<sup>8</sup> It is possible, though unlikely, for someone to accept this latter claim even though she believes people are always morally required to promote a single, overriding aim, like securing the greatest happiness for the greatest number.<sup>9</sup>

I can better explain why the argument rests on this claim if I consider two examples: 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 against harm. Rather, he observes agent-neutral principles that always permit him to do whatever is necessary to produce the best results, even when this involves harming one person to prevent more people from being harmed.

It may be that when someone realizes that Jones is willing to harm her in order to help others, she will be afraid to interact with him. But, given that she has decisive reason to promote the good, it is inappropriate for her to fear him, avoid him, or avoid trusting him with knowledge he might use to harm her. In other words, if people have decisive reason to do whatever is necessary to promote the good, they have decisive reason to sacrifice their interests when this will produce the best results. So, Jones' willingness to harm people in order to promote the good does not,

---

<sup>8</sup> For a recent defense of this claim, see Scheffler 2010.

<sup>9</sup> See Sidgwick 1907: Concluding Chapter.

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 any constraints against harm. In this second scenario, when someone reasonably promotes her own aims, it does not make sense, other things equal, for her to let Jones harm her, whether or not the harm is somehow necessary to produce the best results impersonally considered. In other words, there are cases in which Jones' harming someone in order 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 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 possessions 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 unworthy of their civic trust.

To be clear, Jones is untrustworthy even if his willingness to harm some to save others lowers everyone's risk of being harmed. Of course, if Jones' adherence to his principles lowers others' 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 is better able to help people the more he knows 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 has reason to be discriminating in what she lets Jones learn about her. That is, 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. And even when she has reason to let Jones learn something about her, she also has reason to rely on force or vigilance to ensure that he does not use what he knows in ways that make her worse off. In short, 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.

So, given that it is sometimes reasonable for people to pursue their private aims, a person cannot be worthy of people's civic trust if she is willing to do to them whatever is necessary to produce good results. Rather, to be worthy of such trust, she must observe agent-relative prohibitions against mistreating them. More precisely, she must observe three prohibitions against harm, including two constraints.<sup>10</sup> Together, these prohibitions accommodate what I believe to be people's most important intuitions about constraints against harm.

(1) First, one must observe a prohibition against doing non-optimal harm, or in other words, harm that is not necessary to produce the best available results. This principle prohibits someone from doing harm to others in order to promote her private aims, and from showing callous indifference to harm her actions might cause. To be clear, the principle is not, by itself, a constraint. To the contrary, it can be derived from the Act Consequentialist requirement that we always do whatever will produce the best results. But it is unsurprising that considerations that justify constraints – which prohibit harm that is necessary to promote the good – justify a prohibition against non-optimal harm as well.

---

<sup>10</sup> Each of these prohibitions makes exceptions for certain kinds of harm. I will discuss the main exceptions below.

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 affects her relation to others. Such a person may be willing to tarnish a colleague's reputation to advance her own career, or she may be disposed to hurt people's feelings or damage their belongings out of carelessness. Other things equal, her character gives people reason to stay out of her way, restrain her, or find some other way to prevent her from interacting with them. And when people cannot avoid her, they have reason to worry that they will suffer a net loss as a result. So, if someone is willing to do harm that is not necessary to promote the good, this makes her unworthy of people's civic trust.

(2) Second, to be worthy of civic trust, someone has to observe a constraint against intending harm to other people, whether to promote her 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 harm, she aims to produce the harm, and this aim guides her action 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 sense of self-worth. 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 is likely to embarrass him. She does not intend to cause him distress. Rather, the 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, in

order 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, this person's willingness to intend harm makes her unworthy of civic trust. As she learns more about where people are, what goods they possess, how attentive they are, and so on, she becomes better able to carry out her aim of harming them to promote the good. 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 in order 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 a certain treatment. The doctor will treat the fact that the patient has healthy organs, which can be used to aid others after he dies, as a justification for withholding treatment, and for avoiding behaviors 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 one of the principles someone must observe to be worthy of civic trust enables us to address an important difficulty in formulating that constraint: many cases of intending harm can be

described in such a way that, strictly speaking, the harm is merely foreseen.<sup>11</sup>

Consider the familiar case in which 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 argue that, strictly speaking, she does not intend to kill the child. Rather, she merely intends to crush the child's head, and the death is a foreseen side effect. But if someone can deliberately crush another person's head without intending harm, a constraint against intending harm is not nearly as relevant to moral deliberation 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.<sup>12</sup> But, whether or not we accept this response, the constraint against intending harm that someone must observe to be worthy of civic trust avoids the problem. If someone is willing to crush other 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 these aims. Her willingness to treat people this way makes her unworthy of their civic trust. So, adapting a formulation from Warren Quinn, we might say that to be worthy of civic trust, someone must observe a constraint that not

---

<sup>11</sup> For discussions of this worry, see Bennett 1995: Ch. 11, Foot 2003: 21, McMahan 1994, and Quinn 1993b.

<sup>12</sup> See Foot 2003: 21-22.

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).<sup>13</sup> For ease of expression, I will describe this principle as a constraint against intending harm.

(3) 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 the account I defend, both the distinction between intending and foreseeing harm *and* the distinction between doing and allowing harm are morally significant. This is among the features that make my account distinctive. As Shelly Kagan notes, advocates of the one distinction typically deny that the other distinction has any significance, but this seems misguided, because both distinctions are intuitively important.<sup>14</sup> 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 may be 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 at the result. In other words, the distinction

---

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

<sup>14</sup> Kagan 1989: 87.

between intending and foreseeing harm concerns an agent's aims, but the distinction between doing harm as a causal means and doing harm as a causal side effect concerns the role harm plays in a causal chain leading from the 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 the harm is *unintended*. 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 that she administers to manage this patient's pain might also 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 somewhat aggressive doctor who prescribed the drug, she decides not to pursue the matter, and to keep giving the potentially dangerous drug to the 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 other people 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 the drug are of two sorts, and observing principles of either sort makes someone untrustworthy. On the one hand, the principle might permit someone to do non-optimal harm. I explained above why observing such a principle makes someone untrustworthy. On the other hand, the principle might permit someone to harm people, but only when 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 a certain drug to her 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 the patient 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 facts to justify behavior that would harm them.

Recognizing the link between constraints trust, and appealing to this link both to clarify and to make sense of the content of constraints, enable us to respond to one of the main challenges that accounts of constraints face. Constraints typically prohibit us from doing or intending harm to others, but they also seem to make exceptions, the most important of which concern harm to people who reasonably consent to being harmed, harm in self-defense, and *unintended* harm that is a causal side effect, rather than 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 other people, where this involves – perhaps, among other things – being worthy of people’s civic trust. 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 the relevant sorts of harm to them.

Turning first to consent, a person's willingness to harm someone who has reasonably consented to being harmed does not, by itself, make her untrustworthy. For example, a clerk may sell cigarettes to someone who knows that smoking can cause cancer, but is willing to take the risk; a doctor might perform a preventive mastectomy on a patient who seeks to reduce her considerable risk of developing breast cancer; or someone might engage in limited forms of sexual sadomasochism 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 person reasonably shares. 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, which she might use to carry out 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 the 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 people 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 wants 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 kill him.<sup>15</sup>

Imagine that the bystander turns the trolley, and that causing the lone worker's death is merely a regrettable side effect, not part of her aim. Provided that she observes the three prohibitions I described above, her willingness to harm people in cases like this one, in which 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 tracks at her local station. Because she observes a prohibition against intending harm, they need not worry that she will use what she happens to learn about them to serve her aim of harming them. Finally, since she is willing to harm only as a causal side-effect 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 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 facts to justify behavior that would harm them. To the contrary, each worker – including the one who ends up on the alternate track – has powerful reasons to ensure that the bystander knows facts about his location, his

---

<sup>15</sup> For attempts to describe and defend a constraint against doing harm that permits turning the trolley, see Foot 2003: 23; Kamm 1996: Chs. 6-7 and 2007: Ch. 5; Quinn 1993a; and Thomson 1976, 1985, and 1990: Ch. 7.

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 on a track runs together two facts that, in this context, are importantly distinct. The first fact, namely, that the worker is on the track, concerns that very worker's location and vulnerability, but the second, namely, that there is no one else 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; that is, 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.<sup>16</sup>

---

<sup>16</sup> Though my own intuitions about other, more complicated versions of the Trolley Case are not robust, many people believe that the bystander is permitted to turn the trolley in 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

I should add that, although discussions of this third exception focus 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 direct it to some other 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 other, less densely populated areas (Glover 1977: 102-103).<sup>17</sup> The same considerations that explain why someone who is worthy of civic trust can be willing to harm in the Trolley Case explain why such a person can be willing to harm in these other cases.

In addition to making sense of the main exceptions to constraints against harm, the account I defend enables us to respond to another challenge that faces

---

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-world 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 appeal to this rationale to guide my thinking about such cases.

<sup>17</sup> Warren Quinn (1993a: note 30) describes this case and points out its similarity to the Trolley Case.

views, like mine, that aim to make constraints intelligible by appealing to the moral 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 be 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 behave in ways that make them unworthy of competitors' trust in certain respects. Imagine that Smith lives in a poor community in which many people struggle to find work. Smith is surely permitted to apply for jobs at local businesses and to present himself in the best possible light to potential employers. But his willingness to do so gives others reason to worry that he will take some job they desperately need, and to avoid revealing facts, say, about their work experience, which he might use to gain a competitive advantage.<sup>18</sup>

I said above that being worthy of people's civic trust is a condition for living with them in the kind of harmony that characterizes permissible interaction. But this claim has an important qualification that enables my account to make room for permissible competition. Generally, one has to be worthy of people's civic trust to live in the relevant sort of harmony with them. But in certain forms of competition, people can interact harmoniously, despite being untrustworthy in limited respects, because their outward conflict is a manifestation of some more fundamental unity.

---

<sup>18</sup> People are permitted to be unworthy of competitors' trust in certain respects, not only in cases of competition for scarce resources like jobs or plots of land, but also, for example, in competitive games like poker. Since I do not have space to discuss all such cases here, I will focus on competition for scarce resources, which seems less tractable than other intuitively permissible forms of competition.

Because trustworthiness and harmony can come apart to some degree in such cases, people engaged in these forms of competition are permitted to behave in ways that make them untrustworthy in certain limited respects.

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

Competition, together with certain limited forms of untrustworthiness that may be associated with it, can be unobjectionable on my account if they are aspects of some fair and efficient procedure for distributing a society's resources. For example, 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 of promoting 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 certain limited respects,

but the conflict between Smith and his rivals rests on an underlying harmony that is grounded in their sharing, or having decisive reason to share, a single aim.<sup>19</sup>

These considerations enable my account to accommodate permissions to engage in certain forms of competition, but, to be clear, they do not justify permissions to violate the constraints against harm that I described above. I just argued that a person is permitted to behave 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, say, deliberately harms another person without her 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 in order to promote the good, it is appropriate for her to resist, flee, and so on. So, unlike someone who is merely willing to compete for scarce resources, someone who is willing to harm people without their consent, whether to promote his own aims or the greater good, is untrustworthy in a sense that precludes his living in harmony with people.

#### IV

---

<sup>19</sup> Allan Wood (1999: 169-170) offers a similar defense of the view that Kant's Kingdom of Ends formulation of the Categorical Imperative – which expresses, somewhat differently than my account does, the idea that acting rightly essentially involves living in harmony with others – can accommodate permissible competition.

Now that I have explained what civic trust is and why observing certain constraints is a condition for being worthy of such trust, I can better 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 they pursue reasonable aims by reasonable means. To be clear, this does not imply that someone who acts rightly lives in harmony with people who show callous disregard for others' interests. To the contrary, if someone, say, organizes a boycott to disrupt the operations of a factory whose owners cruelly exploit their workers, her actions may be both morally admirable and fundamentally discordant with the factory owners' aims. Rather, the view states that when someone acts rightly, she thereby lives in a kind of harmony with people who respond, not only to reasons to promote their own interests, but also to reasons to promote or protect others' interests.

This view is a near platitude, accepted in some form or other by a broad range of philosophers. One influential statement of this view is Kant's Kingdom of Ends formulation of morality's ultimate principle. Kant claims 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). He 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). 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 our 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, in part, 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 a person's recognition that she has acted wrongly is her sense that she has disrupted her harmonious relations to others.

Mill's account of moral motivation contains another, importantly different formulation of this view (Mill 1863: Ch. 3). According to Mill, acts are right insofar as they promote happiness, where no one's happiness counts more or less than anyone else's, other things equal. He claims 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 essentially involves living in some kind of harmony with others makes sense 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 because each of us is vulnerable to the influence of countless other people, and none of us 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 complex, difficult aims, often lacking the

resources needed to achieve these aims without help. Furthermore, even if someone manages to survive and achieve her aims more or less on her own, she is unlikely to live a *good* life unless she shares her life with people she cares about. The view that when someone 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 somehow accommodate the judgment that everyone is, in some respect, just one person among others. In other words, 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 or aims matter more 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 limit each person's pursuit of her private aims in ways that bring her life into substantial harmony with others' lives, provided that these others pursue reasonable aims by reasonable means. Limiting each person's behavior this way – as opposed to, say, permitting her to pursue her own interests or the interests of her friends without regard for others who might be affected by her actions – seems necessary for accommodating the judgment that each person is just one among others, impersonally considered.

The view that we live in harmony with others when we observe moral requirements can help provide a rationale for constraints, but only if we understand “harmony” in the right way. On one natural interpretation, which I reject, someone

who acts morally lives in harmony with others by promoting the 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 a person sometimes has sufficient reason to promote her private aims: reason to marry someone she loves, to pursue a career she is passionate about, or merely to take up a hobby she enjoys, even if she 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 such a way that one lives in harmony with others, even when their reasonable aims differ from one's own.

It may be that someone has to satisfy several conditions in order to live in this latter sort of harmony with others. Perhaps she must observe principles that others cannot reasonably reject, observe principles that can be willed into universal law, or be disposed to make considerable personal sacrifices to promote others' interests. I argue that, whatever else she must do to live in such harmony with people, she must be worthy of their civic trust. In particular, when we consider what a person's life in a community is like when she is worthy of civic trust, and what it is like when she is not, it becomes clear 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 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 own aims in ways that make it appropriate for them to pursue *their* reasonable aims openly, without guarding constantly against her learning where they are or what goods they possess, or guarding against her using such knowledge in ways that might hurt them. In short, it makes sense for people around her to adopt attitudes and behaviors that not only make life in a community possible, but also make it worthwhile.

By contrast, someone who is unworthy of such trust might act 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 on a subway platform. Or she might act in ways that make it appropriate for them either 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 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 aims. Rather, it also essentially concerns the quality of people's psychological lives. To live in the relevant sort of harmony with others, a person has to act in ways that do not make it *inappropriate* for them to have certain forms of trust in her – ways that do not make it inappropriate for them to be a part of her community, or to pursue their various projects in an open and unguarded way. Life in any physical community – for example, a neighborhood, a school, or a workplace – that lacks this aspect of

interpersonal harmony is obviously 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 dimension of harmony is impoverished as well.<sup>20</sup> It overlooks considerations that are vital for fully appreciating the appeal of living in harmony with others, and therefore, vital for fully grasping the force of our reason to observe those principles adherence to which helps constitute such harmony.

The claim that someone has to be worthy of people’s civic trust to live in the relevant sort of harmony with them is a pivotal claim in the rationale for constraints that I defend. It supplies a link between the 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 can reasonably pursue certain of their private aims, one cannot be worthy of their civic trust if one is willing to do to them whatever is necessary to produce good results. Rather, to be worthy of people’s civic trust, one must observe certain constraints against doing and intending harm to them, constraints that have the various features I described in Sections I and III. 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 entering into this sort of harmony with them, it follows that moral principles must include constraints that have the features I described.<sup>21</sup>

---

<sup>20</sup> I am indebted to Geoffrey Sayre-McCord for this formulation.

<sup>21</sup> Though he leaves open the possibility that we should observe some form of constraint against harm, Scanlon (2008: Chs. 1-3) denies that this constraint is a prohibition against intending harm, or against

I will close by returning to a topic I raised in the opening section: the relation between the account of constraints that I defend and the accounts in the literature. As I said above, each of the main accounts in the literature has important shortcomings. The account I defend – 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 a way that helps us make better sense of the justification and reason-giving force of constraints, and helps us better understand the deeply plausible conception of morality on which the account rests. To be clear, my aim in this section is not to raise *new* criticisms of accounts in the literature, but rather, to survey some important criticisms that help show that my account supplies an important part of the rationale for constraints.

Some of the accounts in the literature, including Rule Consequentialist accounts, are instrumental; they state that our reason to observe constraints derives from the fact that this is a means of producing desirable results. As I said above, I believe this approach supplies part of the rationale for constraints, but, like many people, I doubt that our obligation to observe constraints depends primarily on the

---

doing harm as a causal means of promoting one's aims. Scanlon's main objection to these formulations of the constraint is that he has not yet seen a satisfactory account of the moral significance of the distinction between intention and foresight, or the distinction between causal means and causal side effects. I have argued, against Scanlon, that we should observe *both* a constraint against intending harm and a constraint against doing harm as a causal means of promoting one's aims. On my view, these constraints derive their significance, in part, from the fact that someone must observe them in order to live with others in the sort of harmony that characterizes permissible interaction.

benefits we would thereby produce. The difference between my approach and the instrumental one is clearest in cases in which people's accepting a permission to maltreat each other in some limited respect would produce better results overall than their accepting constraints against such maltreatment. I have argued that we should observe those constraints adherence to which is a condition for being worthy of civic trust. Because outcomes in which people are trustworthy are, as I noted above, desirable, instrumental accounts can appeal, in their own way, to claims about the link between constraints and trust. But our being trustworthy is just one among many factors that can make outcomes better. Outcomes can also be made better by people's happiness, by people's freedom to direct their own lives, and so on. So, there are cases in which, although people observe constraints and are therefore worthy of civic trust, they fail to produce the best available results. In these 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 so, produce somewhat better results on balance than everyone's accepting a constraint against such detentions. In that case, instrumental accounts imply that such detentions are permitted. By contrast, since our willingness to detain people in this way would make us unworthy of civic trust, my approach implies, plausibly, that these detentions are prohibited. On the view I defend, observing constraints is not just a matter of following principles adherence to which would have, say, 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 special attention to her private interests, each is, in one important sense, just one person among others.

There are also non-instrumental accounts of constraints in the literature, but these accounts, like their instrumental counterparts, are substantially incomplete at best. The most widely discussed of these accounts can be usefully 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 each other is somehow explained by the fact that we have a certain desirable moral status, namely, the status of inviolability, just in case we are so required.<sup>22</sup> But even if we suppose, for the sake of argument, that this curious account succeeds, it seems only to justify the view that we should observe constraints of some sort of other.<sup>23</sup> After all, our being subject to any one of many different sets of constraints would render us inviolable in some respect, and to some degree. So, it seems that we would have to appeal to other considerations to determine whether constraints prohibit people 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.<sup>24</sup>

---

<sup>22</sup> See Quinn 1993a, Kamm 1996: Ch. 10, and Nagel 1995: 83-93. Kamm discusses this moral status in detail, and provides an account of its desirability, in Kamm 2007: Chs. 1, 5, and 8.

<sup>23</sup> For criticisms of this account, see Kagan 1991 and McNaughton and Rawling 1998.

<sup>24</sup> Paul Hurley (2009; Ch. 6) defends another account of constraints that falls in this first category. Hurley's account – which revises and extends Scheffler's (1994: Chs. 2-3) rationale for moral permissions to pursue one's private aims – appeals to claims about the moral significance of the

Kant's moral theory provides the basis of a second category of non-instrumental accounts.<sup>25</sup> 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 to 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.<sup>26</sup> In contrast to this Kantian approach, the account of constraints I defend rests on *substantive* claims about the relation someone bears to people when she acts rightly.<sup>27</sup>

---

personal point of view to provide rationales both for permissions to pursue one's private aims and for constraints. If Hurley's account succeeds, it makes the agent-relativity of constraints more intelligible, but it does not purport to make sense of the more nuanced features of constraints that I described.

<sup>25</sup> My remarks focus on Kant's statement of his views in the *Groundwork of the Metaphysics of Morals*.

<sup>26</sup> For a recent discussion of this criticism, see Scanlon 2011: 117-126.

<sup>27</sup> Arthur Ripstein (2009: Chs. 1-2) develops an illuminating Kantian account of legal constraints, 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 the reason-giving force of moral constraints. Kant's political philosophy rests on his

Finally, Scanlon's Contractualism provides the basis of a third type of non-instrumental account (Scanlon 1998: Chs. 4-5). Contractualism's main claim, which could be developed in a variety of ways, is that someone acts rightly when and because she acts 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: put roughly, 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

---

characterization of the relation people bear to one another when they observe just laws. For Kant, this relation consists solely in people's limiting their outward conduct in ways that prevent them 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 people bear to one another when they observe *moral* constraints focuses not only on the outer form of people's interaction, but also on their inner 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 these constraints is worth caring about – until we recognize that observing these constraints makes it appropriate for people to trust us in certain ways. The account of moral constraints that I defend rests on the view that observing constraints makes us worthy of such trust.

restrictions, Scanlon's view can provide a rationale for constraints.<sup>28</sup> But I believe these restrictions are implausible – claims about the impersonal value of outcomes and claims about the number of people who may be helped or harmed by someone's behavior are sometimes directly relevant to the rightness or wrongness of actions.<sup>29</sup>

Unlike Scanlon's view, and unlike accounts of constraints that are derived from it, the account I defend accommodates the view that claims about impersonal value and claims about numbers are sometimes, by themselves, morally relevant. In other words, my account offers an alternative to Consequentialist thinking about constraints, but not, as Scanlon's view does, by ruling out as intrinsically irrelevant two of the main factors to which Consequentialist reasoning appeals. Rather, by appealing to claims about the relation someone bears to people when she acts rightly, my account justifies limits on the ways these two factors help determine how we should act.

In this 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 the account I defend 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 said throughout, some of these accounts may supply vital parts of the rationale for constraints. Nor, for that matter, do I claim that the comprehensive moral theories from which some of these accounts derive should be rejected entirely. I have argued

---

<sup>28</sup> See Kumar 1999: 304-309. Stephen Darwall claims that "it is well known that ... constraints can be derived within Contractualism", and he attempts to give this approach "a deeper philosophical rationale" (Darwall 2006: 37).

<sup>29</sup> Parfit (2011: Ch. 21) develops this objection to Scanlon's view.

that our obligation to observe certain 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 both 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 do not have space to address this issue here. My aim has been, not to refute the accounts of constraints in the literature or to provide an alternative to the comprehensive theories in the literature, but rather, to shine a light on a certain value – namely, civic trust – that one must appreciate in order to fully understand the justification and reason-giving force of certain constraints.<sup>30</sup> I have argued that recognizing this value enables us both to better understand constraints and to better understand what it means to live with other people in the kind of harmony that, according to a familiar and deeply plausible view, characterizes morally permissible interaction.

## 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

---

<sup>30</sup> I am indebted to Geoffrey Sayre-McCord for this formulation.

Press, 130-151.

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

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

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 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*.

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. Forthcoming. Faith in Humanity. *Philosophy and Phenomenological Research* 87: 664-687.
- 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.
- \_\_\_\_\_. 2008. *Moral Dimensions: Permissibility, Meaning, Blame*. 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.