

Structural Injustice and Individual Responsibility

Andrea Sangiovanni

One often hears about structural injustice. The term is used to indicate, roughly, how entrenched inequalities of status, power, and resources can conspire to create unjust conditions without individuals being either aware of or morally responsible for the injustice. But it is obscure what, more precisely, the term refers to, and what the relationship is between structural injustice and individual responsibility. In this essay, I first identify a series of distinctions that will help us to make sense of the variety of meanings associated with the term. I then defend an account of the conditions under which individuals can be held to account for how they contribute to and reproduce structural injustice. If I am right, then we are individually responsible for a much wider variety of structural injustice than we might have otherwise thought.

I

Structural injustice is injustice somehow related to the existence of a structure. Before we can understand the relation between injustice and structure, however, we need to know what a structure is. Let us begin with what we might call an *abstract* structure, which is an abstract representation of the set of patterned relations possible for a given set of objects.¹ The abstract structure of chess as a game, for example, is given by a list of pieces, a definition of the field of play, and a set of rules governing possible moves within the field for each piece. Analogously, an abstract *social* structure is an abstract representation of the patterned social relations possible for a given set of roles. Courtly life in the Renaissance, for example, has an abstract structure composed of not only the formally stated roles, prohibitions, and permissions defining courtly life but also, as William Sewell puts it, those “not always conscious” informal “conventions, recipes, scenarios, principles of action, and habits of speech”—including the shared concepts, beliefs, assumptions, and other attitudes with which individuals make sense of the social world and their place in it. These rules, concepts, beliefs, assumption, conventions, recipes, scenarios, principles of action, and habits of speech are the stuff of cultural and social anthropology. Sewell aptly calls them *schemas*. They define the positions at court; the rules, conventions, scenarios,² that govern possible moves within courtly life (including what is “out of bounds”); and the character of the social relations which can be constituted by particular applications of the schemas.

But social structures are not just abstract objects. They are enacted by particular people in particular places at particular times with particular bits of the world (textiles for clothing; iron for weapons, and so on.) Abstract social structures, as enacted by particular people in particular times and places then become *practices*. As lived practices, abstract social structures and their constituent schemas both give shape to and are shaped by the *resources* that different actors come to possess within the practices and by the *decisions* they make within them. Resources are, roughly but adequately for our purposes, any material or nonmaterial thing that is useful for getting what we want. Clamshells in the Kula ring are resources, but so is being the king, which gives the king authority and power over his subjects. Schemas, in turn, can both *bestow* value on certain resources, for example, by shaping their social meaning (e.g., clamshells) and can *constitute* their value by establishing what they are in the first place (e.g., kingship). The reproduction of schemas depends crucially on their perceived legitimacy in the eyes of those under their sway, and that legitimacy will often be contingent on the ability of the schemas to meet expectations regarding resource accumulation and distribution.³ Resources and schemas are therefore mutually constitutive: without schemas, many resources would not exist; and without resources, schemas could never be enacted. At the same time—and very importantly when trying to explain how social structures can be transformed through time—schemas can be, in turn, altered, abandoned, recast, and so on, in response to events, circumstances, and human action (e.g., the death of the king, revolution, and so on.)⁴ I therefore adopt the view—shared by Sewell but also by Iris Marion Young and Sally Haslanger⁵—that *social structures simpliciter* (rather than merely *abstract* social structures) are schemas, as enacted in practices by particular people in particular places with particular things, that both are shaped by and give shape to distributions of material and nonmaterial resources.

II

So far I have outlined an account of social structure. It is broad and encompassing, and I hope uncontroversial. What matters is what we do with it, and especially how we construe its relation to justice and injustice. I will distinguish five possible understandings of what the term *structural injustice* refers to. They each draw on the idea of structure in a different way, and are often run together in debate and discussion.⁶ This discussion will allow us to focus more carefully on what is most controversial about Iris Marion Young's project, which has done the most to bring our attention to the importance of structural injustice.

Structural injustice can be either

1. ...the violation of principles of justice that only apply in the presence of interactions mediated by social structures;
2. ...injustice that is a property of structures (but not of individual actions);
3. ...injustice whose criteria are comparative rather than non-comparative, that is one must look at the overall pattern or structure of distribution to determine whether any particular holding is just; one cannot, that is, look merely at individual-level attributes in isolation to determine what each person ought to have according to justice;
4. ...injustice that is explained by the presence of a structure rather than by an aggregate of individual decisions operating independently or in spite of a structure; or
5. ...injustice that is explained as in (4), and where the vast majority of participants have not engaged in any wrongdoing in bringing it about.

Notice that (1)–(3) are claims about the nature of justice itself, whereas (4) and (5) are claims not about the nature of justice, but about what best explains injustice in our world, and what our individual responsibility for such injustice is. John Rawls, for example, endorses (1)–(4).⁷ Young's account of structural injustice, on the other hand, is not, I believe, best seen as aiming to evaluate (1)–(3), which are essential (and controversial) commitments of the Rawlsian project but not, it seems to me, necessary for Young's.⁸ I will return to the importance of distinguishing (1)–(3) from (4) and (5) below. For the moment, I will focus on (4) and (5), which articulate the main claims I will assess in this article.

Explanation is structural in the sense of (4) when it points to the causal role played by a social structure in producing particular patterns of action. Structural explanation is most obvious when we are searching for causal explanations for event *types* (e.g., revolutions in general) rather than *tokens* (the French Revolution). Imagine, for example, that we are trying to explain why there are fewer women at the professorial level in the STEM subjects than men. A structural explanation will point to the role of, say, gendered expectations regarding choice of subjects for girls and women, lack of affordable childcare, absence of shared parenting norms, and implicit bias in the evaluation of academic records and hiring decisions. These are all structural variables because they point to the role of those broader institutions, resources, and societal norms in light of which individuals' beliefs, desires, and other attitudes are formed, rather than directly to the attitudes and actions of particular individuals. While of course such individual-level actions and attitudes are also causally important, a structural explanation will insist that such attitudes and actions alone would be insufficient to explain the gender-based patterning we see in the result.

But structural explanation can also be important in explaining *singular* event tokens. Suppose that Jane is not a professor, but a lecturer. And we

ask: Why is she not a professor? We might point to particular decisions that she took (for example, to take time off to raise the kids, a decision that led her to publish less than many of her peers), or to particular decisions that others took (for example, the panel turned down her promotion because they felt that she was still too junior). And we might do the same for all women in Jane's position. These would be individual-level explanations. But we might also be missing something. If it turns out that this kind of individual-level explanation is recurrent across women but *not* across men, we might say that a fuller explanation must go deeper. We might then wonder: "Yes, but what explains why Jane *and many women like her* chose to stay home? And what explains why many fewer men chose to do so?" What we would then be looking for is not simply an individual-level factor, but a *structural* factor that explains the *pattern* in women's and men's choices—a factor that, in turn, might also help us to explain why *Jane* in particular was not promoted. And, once we shift our focus, it might become clear that Jane (like many others) chose to stay home because *there was no affordable child-care*, because her husband (like many others) was reluctant to stay home *because he felt he would lose respect at work if he did*, and because she reasonably expects to earn less than her husband *given the gender pay gap*.⁹ At this point in our explanation, we would be invoking structural factors (gendered expectations, lack of institutional provision, and so on) to explain the set of choices she faced, which of course also has a causal role in explaining the *particular* choice she made (and the reproduction of such choices and patterns set in the wider society!).

Note, finally, that the account of structural explanation we have given is, for all we have said, evaluatively neutral in the following sense. We might be convinced that the structural variables just mentioned are important in any empirical-causal explanation for Jane's situation, but deny that she suffers any injustice, or that any of the actors involved have engaged in any wrongdoing. This confirms that (4) asserts a primarily causal claim, which leaves open whether particular situations (say Jane's) are unjust, or particular agents wrongdoers.

This leads us to (5). (5) asserts the same causal claim as (4), and so also registers the presence of an injustice caused, in part, by a structure, but adds a further moral evaluation, namely that most individuals have not engaged in wrongdoing in bringing it about. On this view, structural injustice arises from the operation of structures that are, according to Young, "produced and reproduced by large numbers of people acting according to normally accepted rules and practices," but where the harms produced by such compliance are so diffuse that they "cannot be traced to particular contributors to the practice." "Usually," she continues, "we enact these [oppression-producing] conventions and practices in a habitual way, without explicit reflection on the wider implications of what we are doing, having in the foreground of our consciousness and intention [only] those immediate

goals we want to achieve.”¹⁰ Because of the diffuse nature of such reproduction, and the fact that very few individuals intend to propagate injustice or oppression, it looks implausible (except in special cases of express intention, causal impact, and so on) to hold participants morally responsible for it. This is a central thesis of Young’s work on structural injustice. This explains, Young argues, why structural injustice is so often overlooked: if no individual is a wrongdoer, then the temptation is to conclude that there must not be, for that very reason, any injustice at all in situations like Jane’s. Young marshals a similar critique against those who believe that all the oppression suffered by historically disadvantaged groups—women, ethnic and racial minorities, the LGBT community—can be explained solely by theories of discrimination, which are problematically “agent- and fault-oriented concepts,” and hence leave too little room for structural factors.¹¹

I will have much more to say about Young’s account of responsibility and, more generally, about the extent to which individuals can be morally responsible for structural injustice in the next section. In concluding this section, I want to emphasize how important it is to keep (1)–(3) and (4)–(5) separate. Note, for example, that one can hold (4) and (5) and be entirely neutral with respect to the content, scope, or grounds of justice (and so with respect to [1]–[3]). Take Jane, for example. We might believe that the distribution of opportunities that leaves Jane with fewer choices than Jack requires structural explanation, but, at the same time, believe that it is objectionable from a point of view that asserts (a) that principles of equality of opportunity apply among participants in a basic structure only insofar as they apply among individuals as such, as many cosmopolitans do (in opposition to [1]); (b) that the injustice can only be a property of individual choices (in opposition to [2]); and (c) that Jane deserves to have the same opportunities as men of similar talent on purely non-comparative grounds (in opposition to [3]). We might, that is, hold that structures are causally important in explaining injustice *without* endorsing the presence of structural injustice as described by (1)–(3).

It is also important to keep (4) and (5) separate. If one doesn’t, it leaves one wondering whether there can be structural injustice where most participants *are* wrongdoers: if there are such wrongdoers, then does this mean that the injustices in question cease to be structural? It strikes me as clearer to claim that structural injustice is injustice according to (4), and leave the question of moral responsibility to independent argument, to which we now turn.

III

So far, I have elucidated the concept of a social structure, distinguished five different claims associated with structural injustice, and suggested why drawing these distinctions is important. As I mentioned at the end of the

previous section, I believe we can best explore the issues raised by (4) and (5) by disentangling the idea of a structural explanation (as in [4]) from the separate question—What kind of moral responsibility, if any, might those who participate in reproducing injustice-causing social structures bear (as in [5])? The latter issue raises notoriously difficult challenges regarding the role of imperceptibly small causal effects, social meanings, and the possibility of change through political action. I will argue that Young’s approach to this issue is unsuccessful. I will then clarify the conditions under which “everyday” participants in the reproduction of injustice-causing social structures bear backward-looking moral responsibility for wrongdoing.

It seems obvious that many injustices of the kind experienced by Jane in the example above arise as a result of myriad individual actions, adjustments, choices, and patterns of institutionalized and uninstitutionalized behavior that are not objectionable if taken *one by one*. But why, given the very great harm that such actions do when taken *together*, does Young believe that contributing individuals are not wrongdoers? As we have already mentioned in passing, there are three sets of reasons that Young advances for this conclusion. First, very few of those involved (let us assume for the sake of argument) acts with the intention of, in our case, leaving women worse off than men. When Jane’s husband argues that it makes sense for him to go to work (given the larger pay he is likely to receive and given the reputational costs he will suffer if he does not) and when she agrees, neither acts with the intention of reproducing the pay gap between women and men. And, more subtly, very few involved act with the intention of playing their part in achieving a larger collective goal of leaving women with less than men (as, say, the Dresden bombers each played their part in the collective action of bombing the city).¹²

Second, each individual’s causal contribution¹³ is not, Young asserts, *traceable* to any particular individual. Such contributions are not traceable to any particular individual both because each individual’s contribution is not enough, on its own, either to give rise to or to dismantle a social structure and because it is not possible to connect any particular aspect of the social structure’s existence or persistence to any individual’s causal contribution.¹⁴ Whether or not, that is, Jane’s employer, her husband, Jane, or any other individual person had acted differently (holding constant the actions of everyone else) would not have made any perceptible causal difference to the larger pattern of inequality between women and men, and no one person’s actions are causally responsible for any specific aspect of the gendered social structure. In the language of lawyers, no one is a “but-for” cause of the inequality. Young argues that, if this is true (and if it is also true that no individual intended to reproduce the inequality in acting in various structure-complying ways), then it looks implausible to argue that any such individual had, in reproducing the injustice, engaged in any wrongdoing.

Third, Young argues that holding individuals responsible is strategically misguided because “it tends to make people defensive and to engage in ‘blame switching.’”¹⁵ Trying to find a culprit, a person on whom to pin the wrongdoing, makes the project of fighting injustice more difficult to carry out because people accused in this way are likely to turn defensive, and hence less likely to engage in shared political struggle.

Instead, Young argues that we need to shift perspective:

The lesson I think we should take from my objections . . . is that responsibility in relation to structural injustice should not be thought of as an attenuated form of [backward-looking] responsibility as complicity. . . . What we should seek is not variation on a weaker form of liability, but rather a different conception of [forward-looking] responsibility altogether. . . . The social connection model finds that all those who contribute by their actions to structural processes with some unjust outcomes share responsibility for the injustice. *This responsibility is not primarily backward-looking, as the attribution of guilt or fault is, but rather primarily forward-looking.* Being responsible in relation to structural injustice means that one has an obligation to join with others who share that responsibility in order to transform the structural processes to make their outcomes less unjust.¹⁶

It is essential to the idea of a “forward-looking” responsibility that the duty is not grounded in the fact one is morally responsible for wrongdoing. The idea of a “forward-looking responsibility” here is akin to the idea of a role responsibility, such as a duty of office, or the duties of lifeguards, or, indeed, the duty of a bystander who is causally involved in a crime for which they are not morally responsible. Suppose, for example, that you are in a stadium in which there is a terrorist attack; many are injured but you are not. You are causally involved insofar as the terrorist would not have attacked had there not been so many people in the stadium (of which you are one). Assume, furthermore, that you have a duty to aid those who have been injured. Whatever the basis of this duty (on which more below), it is not based on your being a wrongdoer; it is, in this sense, entirely forward-looking. In the same way, we are said to bear a forward-looking responsibility to stop structural injustice through collective action because of our causal involvement in its reproduction.¹⁷

Young’s argument for switching from backward- to forward-looking responsibility strikes me as unconvincing. Let us take Young’s third reason for invoking forward-looking role responsibilities first, namely that this will have beneficial effects on the likelihood that people will take up action to remedy injustice. And let us assume that Young, on an empirical level, is right: insisting on backward-looking responsibility will make progressive reform less likely, while advocating forward-looking responsibilities will make it more likely. Even in this case, we want to know: Whatever the effects on motivation and whatever we might want to advocate in public or for

strategic reasons, do individuals—as a moral matter—lack backward-looking responsibility or not? This further question, which is the one we are interested in, cannot be answered by appealing to strategic considerations.

What about Young's first and second reasons for "switching" to a forward-looking conception of responsibility? There are two problems. First, it is obscure why causal involvement in an injustice should matter if *not* because it implicates the actors in a backward-looking sense. Returning to our terrorism case, to explain your duty to aid, why does it matter that you are causally involved? Wouldn't the same duty apply to someone who was in your same position, but who only entered the stadium *after* the terrorist decided to attack? In the bystander case, the fact that someone was causally involved seems to matter only insofar as it makes it more likely for that person to be able to do something to help; this is not, however, a necessary condition for the duty to exist. Indeed, holding constant their ability to help (proximity, medical knowledge, and so on), there is no more reason for someone who was *merely* causally involved (without being morally implicated) to help than someone who wasn't.

The second problem emerges when we consider an amendment to Young's view that seeks to rescue the idea that causal involvement matters morally for non-instrumental reasons, that is, that matters not simply by putting the actor in a better place to help. To see it, one needs to distinguish between liability, moral responsibility, and culpability (in a way Young never does in *Responsibility for Justice*). I might be morally responsible for wrongdoing, say, by punching Abe, by which we mean that the action is both wrongful and ascribable to me. But I might not be liable for damages or blameworthy if, for example, someone was threatening to kill my children in the event I didn't punch Abe. So someone can be morally responsible in this sense without being liable or culpable; we have an *excuse* but not a *justification* for our punching Abe. In the language of lawyers, I have a defense of duress. This kind of case can be contrasted with a case in which I suffer from a spasm that results in the same injury to Abe; in the latter case, I am not morally responsible because the action of punching Abe is not ascribable to me as a doing or a trying (and so of course I am not liable or culpable either). There was no battery and hence no wrongdoing at all.

With this tripartite distinction in the background, one might revise Young's account in the following way. Young could say that individuals causally involved in the reproduction of injustice are indeed wrongdoers (and hence morally responsible in the sense I just mentioned); however, because the contributions to injustice are unintended, tiny, within normally accepted rules and patterns of action, and not traceable, individuals are neither culpable nor liable for their wrongdoing. Most "everyday" wrongdoers, that is, have an excuse; despite this excuse, their being wrongdoers explains why they have remedial duties they would not have otherwise had as mere causally involved bystanders. This revision preserves something from

Young's account, namely that, despite the fact that we are not liable for damages or blameworthy for our participation in injustice, we have special, remedial duties to address injustice; however, it rejects the idea that individual responsibility for structural injustice is not backward-looking (since it requires us to accept that individuals are wrongdoers, albeit wrongdoers with an excuse). I believe we ought to accept something like this revision as the best account of the relationship between individual responsibility and structural injustice. But to do so, two questions remain to be answered, neither of which was addressed by Young, but which are crucial. First, we need an account of why, if individual contributions to structural injustice are unintended, tiny, within normally accepted rules and patterns of action, and not traceable, individuals are wrongdoers at all. Even if, that is, we accept that everyday participants in structural injustice are neither liable nor culpable, why is our everyday reproduction of structural injustice *not* more like the terrorist case than the case of duress? Second, we need to identify whether the untraceable, unintended, tiny, and everyday character of our contributions really does exculpate in the way our revision of Young's view suggests.

IV

As will become evident in a moment, the best way of querying the moral role of untraceable, unintended, tiny, and everyday contributions is by drawing an analogy to a modified version of Derek Parfit's example of the Harmless Torturers.¹⁸ Suppose that the electrical systems of 1,000 households are wired to a torture machine. If none of the lights in the houses are switched on, then no current runs through the machine. If 1,000 lights are turned on, the person will be in severe pain. But any one light switch makes at most a barely perceptible¹⁹ (and in many cases an imperceptible) difference to the pain experienced by the man connected to the machine, and no one extra quantum of pain felt by the man is traceable to any specific household (we might suppose that the wiring is on timers that make it difficult to say who is causally responsible for which particular effects). Also imagine that each household, in turning on their lights, will only ever deliver one such tiny charge and that the charge will continue to be transmitted as long as they fail to rewire the electrical circuit. And also assume that each household cannot forgo turning on the lights in their house, and that rewiring the circuits would cost them \$1,000. What should we say? Should we say that none of them is morally responsible for torturing the man when they turn on their lights without rewiring? No one individual's action makes more than a barely perceptible difference to the man's pain, and most no perceptible difference at all; furthermore, that difference, if any, is also not traceable to any particular individual. If that is true, how can it be wrong for them not to rewire their house given the considerable costs they would

have to bear otherwise? If we agree with Young, then we must also come to the conclusion that no one is (backward-looking) responsible for the man's torture. And yet, I am suggesting, this is absurd. But how can we explain why it is absurd?

Christopher Kutz offers a powerful, but ultimately unconvincing, proposal to deal with cases like this.²⁰ Referring not to the Harmless Torturers but to global warming (a distinction that will become important in a moment), Kutz argues that the basis for individual moral accountability is a "shared way of life." Sharing a way of life requires much less than participating in a collective project. In collective projects, agents each intend to "do their part" in achieving a shared goal. Kutz uses the bombing of Dresden as an example. The Dresden bombers share a goal, namely to bomb the city, and a set of intentions to promote that goal by each doing their part in the plan required to achieve it. A shared way of life lacks this unified, goal-directed character. While agents who share a way of life share a set of norms, beliefs, desires, expectations, and conventions—put in the terms we presented above, individuals share a way of life when they share a *social structure*—they do not participate in the achievement of an end shared by all. For example, the group of all Anglophones across the world share in the reproduction of a structural schema, namely the English language. But, in speaking English, it would be absurd to argue that Anglophones therefore constitute a collective actor. For Kutz, this is because they do not think of their speaking English as "doing their part" in the realization of a jointly held goal or set of goals.

But if sharing a way of life lacks the intentional character of collective actors like the Dresden bombers, then how can Kutz claim that each individual shares what he calls "inclusive accountability"²¹ for the actions of the group? Kutz argues that once agents begin to *identify* themselves as participants in the reproduction of a social structure, they can also come to see their participation in the shared way of life *as if* it were participation in a collective endeavor. According to Kutz, "even in the absence of a discrete identifiable collective act, agents can think of themselves as participants in a collective venture." In thinking in this way, "drivers can come to be aware of the damage done by a way of life that ignores atmospheric effects."²² They thus become individually accountable for "the unintended consequences of what they do together" just as they are accountable in cases like the Dresden bombers.²³ For this reason, Kutz says that the basis for accountability is *quasi-participatory*—quasi because it is not participatory in the fully intentional and agential sense, but participatory nonetheless.

There are two problems. First, for Kutz, the basis for individual moral accountability in cases of jointly intentional action is the possession, by each participant, of *an intention to promote a collective end by doing one's part*. But if that is so, then it seems puzzling how the account can be extended to nonintentional contexts like global warming (and, indeed, the Harmless

Torturers). In those cases, the very element that explains one's accountability in the intentional cases is *missing*. Agents do not drive their cars in order to do their part in global warming just as the Harmless Torturers do not flip the switches and collect the money in order to do their part in torturing the man. We can draw an analogy (due to Michael Bratman²⁴). When I run the road race, I can *foresee* with complete certainty that I will be wearing down the tread on my trainers; I do not, however, I run the race *with the intention of* wearing down my trainers. I run the race with the intention of winning. Similar things can be said with respect to both the Harmless Torturers and the polluters: I flip the switch with the intention of turning on the lights; I drive my car with the intention of getting to work. But if this is right, then why *should* the polluters or the Torturers think of themselves "as if" they were participating in a collective project of torture/pollution when they are not *in fact and according to Kutz's own analysis* participating in such a project? Why should identification be sufficient? (And what if they didn't "identify" with the way of life that they share in reproducing?) While holding the (false) belief that we are participating in something like a collective project (even when we are not) might have good effects, if sincerely endorsed, its good effects can't justify holding the belief to be *true*.

The second problem is that it is not clear how sharing a way of life in the relevant sense plays an independent explanatory role. Compare, for example, the Harmless Torturers to the global warming case. Although the Harmless Torturers know of each other's existence, and know the consequences of their acts, their decision not to rewire is not an instance of compliance with a structural schema. Unlike the global warming case, the Harmless Torturers (we can easily suppose) do not "share a way of life," and yet they seem just as accountable as the global warmers, and for the same reasons. Imagine, for example, that the tortured man demands justification from the house-owners for what they have done to him; it would be absurd for them to respond: "Oh, but we weren't complying with a structural schema or way of life in not rewiring, so you have no complaint against us." Their accountability seems traceable to their individual actions in perpetrating the collective harm, not in doing so via a social structure. I will return to this point below.

Kutz also argues that our myriad, small-scale contributions to global warming send a demeaning message to those most harmed by it. "Agents," Kutz writes, "can have reason to refrain from participating in a harm, not because of a [causal] relation between this choice and an actual outcome, but because of what the choice symbolizes in their characters and commitments."²⁵ This is a powerful argument but, for it to have force, it presupposes an answer to the very question we are asking: Do agents bear special responsibility for the harms they together aid in bringing about? If the answer is no, then no demeaning message is sent. If the answer is yes, then the expressive meaning of their actions, while adding an extra dimension to the

wrongness of the contribution, does not tell us in virtue of what the contribution is wrong in the first place. And for that we need to revert to Kutz's first argument, which, as we have just seen, is unconvincing.

V

In this section, I will argue both that we can bear backward-looking responsibility for wrongdoing in cases like the Harmless Torturers, and in the next that this account can be extended to illuminate our analysis of structural injustice.

Let me begin with the Self-Torturer's Paradox, which is a single-person case of the Harmless Torturers as I have described it above.²⁶ Imagine you have been made the following offer. You will wear a device that delivers electric shocks in tiny but constant increments of voltage. For each turn of the dial on the device you will receive \$1,000. For each turn of the dial, the increase in any discomfort/pain you feel will be at most barely perceptible (and, in most cases, imperceptible). At setting 1, for example, you barely notice, if at all, the difference from setting 0, and so on for the whole series to 1,000. However, you are also instructed that by the time you reach setting 1,000 on the dial, you will certainly be in unbearable pain. The dilemma is the following. It seems that, for each turn of the dial, you should turn the dial at least once more. After all, as a result of any one turn of the dial, you know that you will experience at worst a barely perceptible difference in pain (or, in many cases, no perceptible difference at all), while receiving, each time, a very significant sum of money. This will be true, furthermore, even at very significant levels of pain. Once you reach those, you can still say to yourself: "Well, but one more turn of the dial will make at most a tiny, barely perceptible difference to the pain I am currently feeling, and in many cases no perceptible difference at all, so I might as well turn." The puzzle is that, if you continue reasoning in this way, you will end up in excruciating, unbearable pain—so severe that, in fact, you would gladly return all of the money you received and then some to return to the initial setting.

It seems obvious that there has been some mistake in your reasoning, given the intransitivity of your preferences. You prefer, that is, setting 1 to setting 0, setting 2 to setting 1, and so on all the way to some very high setting, say setting 950, but setting 0 to setting 950. If you continued reasoning in this way, the operator of the device could pump money out of you *ad infinitum* by repeating the sequence over and over. Given your concern for yourself over time, this would be highly imprudential, leaving you much worse off than you would ever want to be. So what is your mistake, exactly?

Many solutions to the paradox have been offered.²⁷ I will not provide a survey. Instead, I will simply reconstruct the one I believe to be most convincing, extend that solution to the Harmless Torturers, and then, in the final section, extend it again to cases of structural injustice. The solution is

due to Alex Voorhoeve and Ken Binmore.²⁸ The first step in the solution is to realize that it is impossible for the difference in pain between *each and every* turn of the dial to be imperceptible.²⁹ If that were true, then there would be no way for you to be in unbearable pain at setting 1,000. This is why we must stipulate that there are many turns of the dial that will register at least a *barely perceptible* difference (leaving open the possibility that some, but not all, will not be perceptible)—just as we did in the Harmless Torturers. If we define “unbearable pain” as the point at which you would return all the money and then some to go back to setting 0, there must be at least one turn of the dial—or one set of >1 turns—such that you were not in unbearable pain on turns smaller than the smallest member of the set, and you are in unbearable pain on settings that are larger than the largest member of the set.

Assuming that you can play around with the device before starting the official sequence—for example, turning the notch up to various levels, recording the pain felt at those levels in qualitative terms, and then returning it to 0—the next step in the solution is to ask yourself how much money you would require to leave you indifferent between a given level on the dial and a 0 setting.³⁰ Suppose, for example, that you would require \$3,000 to leave you indifferent between 0 and 25 given the degree of pain you expect to feel at level 25 (which you establish by [a] turning up the notch several times to levels at and around 25 and back again, [b] sketching a probability distribution of the pains typical of being at level 25 on the basis of the frequencies reported in [b], and [c] repeating the same procedure for each turn of the dial up to levels at which you require more to remain indifferent than you will receive in reward for those levels). Because you would receive a reward of \$25,000 for turning up to level 25, your overall “gain” for that level is therefore \$22,000. If we assume that there is some level at which you would rather pay back all the money and then some, and we assume that the amount you are willing to pay to remain indifferent between any given level and 0 increases monotonically (the pain never decreases as you increase the voltage), this entails that the indifference curve must remain convex as you increase the voltage from 0 (otherwise there would never be a point where you would pay back all the money and then some). Given these assumptions, there must also be some determinate level where your gain will be maximized. For example, if, at level 53, your expected net gain is \$33,550; at level 54, it is \$32,571; and at level 55 it is \$32,566, you should stop at level 54 since you know that as you increase the levels past 55 your expected gain will steadily decrease until it goes negative (the point of “unbearability”).³¹ Level 54 would then be the level you should choose when the official trial begins. Making the same assumptions about aversion to the pains at different levels, and the existence of a point of unbearability, such a level must exist, furthermore, even if we increase the amounts received at each increment (assuming that this amount is not so large as to eliminate the point

between 0 and 1,000 at which we would rather return the money and then some) or make the increase in voltage between each step smaller and smaller (by extending the scale from 0 to 1,000 to, say, 0 to 100,000).

So where is the mistake in your initial reasoning? The mistake is in your assumption that, from the point of view of the level of pain you expect to feel at level n , it must be the case that an extra \$1,000 could compensate for the extra pain you expect to feel at $n + 1$ *at every point between 0 and 1,000*. Recall that as long as we assume that there is some point at which the pain becomes unbearable before we reach level 1,000, and we assume that the amount you require to remain indifferent between 0 and level n traces a convex and monotonically increasing route as the voltage increases from 0 to n , then there must be a determinate level at which you would maximize your expected return by stopping. The only way you could be fooled into thinking otherwise is if your ability to distinguish the pain between adjacent levels *when presented in sequence* does not give a reliable estimate of the difference in pain you can actually expect to experience at that level.³² But the Self-Torturer shouldn't be fooled: he should realize his tendency to underestimate the importance of small changes and trust rather in his sampling technique, which is more likely to give an accurate estimate of his expected level of pain at different voltages. If he is allowed to sample in the way indicated above, then he can increase the accuracy of his estimates for the aversion he will feel at different levels, and hence also increase his chances of maximizing his return. But even if he is not able to sample, using this reasoning he should come to the realization that there must be some determinate point at which it is no longer wise to turn the dial (even if he cannot determine reliably what that point is).

As the reader will have already noticed, there is an important structural similarity between the Harmless Torturers and the Self-Torturer. But there are also two main differences. First, instead of a succession of 1,000 temporal parts of a single person, each of whom faces a separate decision to turn up the dial, there are a succession of 1,000 distinct selves, each of whom faces a separate decision to turn the dial. Second, instead of inflicting pain on a distinct self, the Self-Torturer's temporal parts inflict pain on themselves taken as a unified whole. These differences are significant—particularly since they involve harming another person rather than oneself—but, as I will now argue, they can be narrowed.³³

Let us first suppose that, for the man tortured in the Harmless Torturers, we have available the probability distributions of his pain reports for each turning of a switch between 0 and 1,000 switches (similar to the probability distributions produced by the Self-Torturer). For example, if 700 people turn the switches on in their houses, then there will be a probability distribution over possible, reportable levels of pain felt by the man at that level of voltage. As discussed above, these probability distributions, while significantly overlapping at adjacent levels, grow less overlapping with levels of

voltage that are further apart. Now assume (as seems plausible) that there is a function that defines how much cumulative gain to the Torturers would be required to leave one *morally* indifferent between leaving the man in the expected discomfort/pain experienced at a given voltage level and leaving the man as he was when the voltage was 0. The function might give us a value of, say, \$3,000 for smaller levels of voltage (at which the man can be expected to feel no discomfort, with a small probability of very mild discomfort). This would mean that anything less than a \$3,000 cumulative gain would not (morally) offset the expected discomfort felt by the man but anything equal to or more than \$3,000 would (assuming linear utilities and interpersonal comparability³⁴). And suppose that all the Torturers agree that this function represents what is morally at stake. If we further assume that there are a series of points (at lower voltages) where the cumulative savings from rewiring will be equal to or greater than the value of the indifference function at that voltage, then we also know that there must be a certain determinate number of light switches allowable which will maximize, from an impersonal point of view, the overall gain (subject to the moral constraints expressed by the function).

Suppose, for example, that we would require a cumulative gain of \$3,000 to leave us indifferent between leaving the man at 0 (without the \$3,000) and at the level of expected (mild to no) discomfort he would experience at voltage level 25. Because we would receive a cumulative gain of \$25,000 for allowing 25 houses not to rewire, our overall “gain” for that level is therefore \$22,000. Given that there is some voltage level at which the man’s expected pain will no longer be offsettable by the cumulative gain to the house owners at that level, and given that the indifference function is convex and monotonically increasing as you increase the voltage from 0, proceeding in this way entails that there must be some determinate number of light switches where the cumulative gain, subject to the moral constraints expressed by the indifference function, will be maximized. Below that level, every further switch leads to a marginally larger net gain; after that level, every further switch leads to a marginally smaller net gain up until the point at which the cumulative gain is not sufficient to morally offset the pain/discomfort felt by the man. After that further point (call it the point of “moral unbearability,” or *m*), not only will failing to rewire not be morally justifiable, but it will be increasingly less justifiable as the numbers of others failing to rewire increases (given that we will judge that increasingly more, relative to the constant gain in savings, would be required to offset the man’s pain as the voltage increases). Notice that this procedure for finding how many switches are morally permissible is exactly analogous to the one we imagined in the Self-Torturer case.

If this is right, what follows for what any house owner ought to do? Ought she to rewire or not? Unlike the sequential, single-person case imagined by the Self-Torturer, it all depends on what she expects others will do

(and how many others there are). If she expects that fewer than $m - 1$ households will fail to rewire (and hence turn on their lights), then she should feel free to turn on her lights without rewiring (discounting by the probability that $x < m - 1$ will switch).³⁵ If she expects that more than m households will fail to rewire (discounting by the probability that $x > m$ will switch), then she should rewire (given that, otherwise, one's light switching alone will produce a marginal [moral] loss). For example, if she expects with certainty that, say, 100 others will fail to rewire, thus transmitting volts at level 100 to the man, and if it would take a \$1,000,000 gain to the household owners to make one morally indifferent between leaving the man in the expected discomfort/pain he can expect to experience at level 100 and not transmitting any current at all, then how should she reason? Given that we are beyond the point of "moral unbearability" (notice that the \$100,000 the household owners would gain if 100 failed to rewire is nowhere near enough to offset the discomfort/pain valued at \$1,000,000), and that the indifference curve is convex (and, let us suppose, *strictly* monotonically increasing across this interval³⁶), leaving the man in the further pain/discomfort he would feel, on average, at level 101 (were I to switch on my lights) must have a greater disvalue (probably much greater) than the \$1,000 she would gain on the margin.³⁷ So her marginal gain, at this point, would not be enough to offset the disvalue of the expected extra quantum of pain experienced by the man. It is for this reason that she ought to rewire.

The thought that I do no wrong in flipping the light switch is therefore only sometimes true.³⁸ If few others flip their switches, then it is true; if many do, then it is false. It all depends on how many I expect will also flip their light switches. In cases where enough others do, and the man is therefore in significant, noncompensated pain, I engage in wrongdoing by also flipping my switch instead of rewiring, despite how miniscule my contribution is. Indeed, this reasoning also applies to anyone who was among the first to turn on their light switch. Once the point of moral unbearability has been passed, they too acquire a moral duty, all else equal, to rewire until they can expect only at most $m - 1$ houses will turn on their lights. This is precisely because their rewiring will produce, given the reasonable assumptions we have made, a larger marginal net benefit from the point of view of all involved.³⁹ Notice further that this is true even if no one quantum of pain/discomfort felt by the man is "traceable" to me specifically; as long as I know that I make a direct causal contribution to the voltage and that my rewiring will produce, on average, a marginal gain (given the indifference function), it does not matter which quantum of pain or relief I am responsible for (if any). Indeed, it doesn't even matter whether my contribution of voltage will lead to any difference in discomfort at all. What matters for our moral evaluation is the *expected* gain/loss. We care what I can reasonably expect *ex ante*, not whether my increasing (or reducing) the voltage *actually* turns out to make a difference *ex post*.⁴⁰ An analogy is helpful: when I drive

my car, or board a flight, I know that I make a causal contribution to the level of carbon in the atmosphere, but I do not know how my particular carbon emission will affect, if at all, global warming (it might, for all I know, be absorbed by a plant that would otherwise die). What matters for determining whether I am engaged in wrongdoing, all else equal, is the *expected* effect of carbon emissions of this kind, including whether there is a marginal gain or loss using an equivalent indifference curve.

VI

We are now in a position to draw the analogy to structural injustice and to compare the solution I have defended to both Kutz and Young's official accounts. Many stand to gain by complying with and reproducing social structures that reproduce injustice. In the gender pay gap example, for instance, in many cases (although certainly not in all), the Jims and Janes stand to gain by choosing Jane to stay home with the kids and Jim to work, employers stand to gain by only paying as much as the market requires, and taxpayers stand to gain by limiting their tax bill. This gain is equivalent, on our analogy, to the \$1,000 saved by our household owners. The result of such individual choices, however, will be the reinforcement of a social structure—constituted by gendered norms and expectations, patterns of often unintentional discrimination, and so on—that leaves women unjustifiably worse off. This is equivalent to the households' maintaining voltage levels beyond *m*. If this is the case, then, by analogy, Jim, Jane, the employers, voters, and so on, all bear backward-looking moral responsibility for reproducing the gender pay gap because their individual choices, given the expected choices of others, *do* make a morally significant difference to the reproduction of injustice. It makes no more sense to say, as Young's account urges us to do—"My causal contribution to the structures reproducing the gender pay gap make either no or very little perceptible difference to the patterns we see; because my ceasing to act in a structure-compliant way would not therefore reduce, to any appreciable extent, the injustice suffered by women, I am not engaged in any wrongdoing"—than for the Harmless Torturers to say—"My causal contribution to the voltage communicated to the man makes either no or very little perceptible difference to the pain he experiences; because my rewiring would not therefore reduce, to any appreciable extent, the pain felt by the man, I am not engaged in any wrongdoing." I have also argued that it does not matter, contra Young, if any particular aspect of the gendered structure is "traceable" to the causal contributions of specific individuals. As long as individuals make a causal contribution to sustaining the unjust structure, the fact that it is difficult to determine which particular aspect of the structure or its resultant injustice is sustained by their individual contribution does not, merely as a result of that fact, mitigate their wrongdoing (leaving aside, for the moment, their

culpability or blameworthiness). What matters is whether their causal contribution makes an *expected* marginal moral difference given what others can be expected to do.

This allows us to return to the second question we raised above: If we are responsible for wrongdoing in reproducing structural injustice, then under what circumstances might we have an excuse—an excuse, that is, that would absolve us of both liability and culpability for the wrongdoing? Of course, any person’s particular situation and particular relation to the social structures they aid in reproducing—where it is known—will change the character of our overall assessment of their degree of wrongdoing as well as their degree of culpability and liability.⁴¹ In many situations, for example, participants will be entirely (and nonculpably) ignorant of how their contributions have an impact. When this is the case, it seems plausible to say that they bear no liability or blameworthiness for their participation, but will as soon as they find out. In other cases, it will be plausible to argue that some are already unfairly burdened when compared to those who are in a better position to remedy the injustice (this might be the case for the victims of injustice themselves). In such cases, it might be appropriate for individuals to wait on others to take action (in the way the Harmless Torturers might in waiting for others to reduce the level of participation to $m - 1$). In still other cases, some will be able to argue that they have compensated for the wrong they perpetuate, not by ceasing to contribute to the wrong (which may be too difficult or very costly), but by taking remedial action in other ways. Two important points follow from these examples of how liability, culpability, and wrongdoing are related and in some cases mitigated. First, in all such cases, individual participation in injustice-producing social structures is in any case participation in (sometimes excusable) wrongdoing, not a mere opportunity to rectify injustices for which we bear no (backward-looking) moral responsibility. Second, the fact that a contribution is tiny, unintended, untraceable, and everyday is not enough *on its own* either to exculpate or to eliminate wrongdoing; while it may reduce one’s degree of wrongdoing (compared with, say, a causal contribution that is much greater or intentional), it doesn’t in itself extinguish liability or blameworthiness. On this account, remedial duties are triggered as a result of wrongdoing rather than as a result of purely “forward-looking” considerations.

One might wonder if this account of possible mitigating factors allows for no one individual or collective agent to be responsible for, say, the gender pay gap *as such*.⁴² In brief, *yes*. The Harmless Torturers is again useful as an analogy here. In that case, we can say that no one individual is morally responsible for the man’s torture *as such*; rather, they are each responsible for *contributing to* his torture. This makes a difference, I have argued, to the degree of wrongdoing in which they are involved (and hence their degree of liability and/or blameworthiness should they lack an excuse). I leave it open whether we also want to ascribe some kind of aggregate or even collective

responsibility to the *group* of individuals for the torture/gender pay gap as such (though, if we do so, we should be careful not to describe the group as a collective *agent* since the group does not exercise the kind of unified agency typical of, say, corporations, or other collectives).⁴³

So how does the account defended here compare with Kutz? Recall that for Kutz we wrongfully reproduce injustice when that injustice is a product of a shared way of life (which, I argued, is equivalent to the idea of sharing a social structure). In those circumstances, we ought to think of ourselves as if we were participants in a collective project (even though we are not in fact such participants). And, just like participants in such collective projects, we ought to think of our responsibility for wrongdoing as traceable not primarily to our causal involvement but to our active participation in the shared way of life. While we may not intend to play our part in the wrongdoing caused by the ways of life we reproduce (as we intend to play our parts in collective projects), we do *identify* with those ways of life. And such identification is enough, Kutz concludes, to make us morally responsible for the effects of our shared way of life. I argued that this was unconvincing: Why would identification without intention be sufficient to make one responsible for wrongdoing? I also claimed that Kutz fails to explain why the Harmless Torturers would be doing anything wrong. After all, the Harmless Torturers do not share a way of life in any morally relevant sense of that term and yet their participation in the torture seems just as objectionable, and for the same reasons, as the global warmers. It should be clear by now that the account I have defended does better: it explains when and why individual participants are morally responsible without appealing to an analogy with intentional, collective projects, makes no use of the idea of “identification,” and easily extends to the Harmless Torturers.

Notes

¹ I draw my account of a social structure primarily from William Sewell (2017), who is reworking and synthesizing Anthony Giddens and Pierre Bourdieu’s use of the term. The account, although different in some nuances (e.g., the relation between practices and structures and the expressive account of social meanings), is compatible with both how the term is used by Sally Haslanger (2012, 414–17) and Iris Marion Young (1977, 60–61), who both also make use of Sewell.

² See de Sousa (1990) for an illuminating account of emotions as narratively structured “scenarios.”

³ Cf. Moore (1978).

⁴ Sewell argues convincingly that including resources is necessary to counteract the more rigid formulations of social structure coming from both Claude Levi-Strauss and Giddens. He also provides a useful account of how change is possible within structures (citing, among other things, the multiplicity and variety of structures shared by a social group, the fact that [abstract] schema can be transposed to many contexts beyond their initial origin, the unpredictability of resource accumulation, the multiple meanings that resources can embody [which often depends on the perspective from which they are viewed]) (Sewell 2017, 140–43).

⁵ See Haslanger (2012, 414–17); Young (1977, 51–52).

⁶ In a chapter entitled “Structure as the Subject of Justice,” Young, for example, sometimes refers to structural injustice as a particular kind of moral wrong (rather than as a particular way in which a wrong is produced) (see, e.g., Young 1977, 45), sometimes as a causal factor in explaining injustice (see, e.g., Young 1977, 70–71, and perhaps most often as an injustice where no one individual is a wrongdoer (see, e.g., Young 1977, 47).

⁷ Re: (1), the two principles only apply, Rawls says, to basic structures, and, as he later clarifies, to the basic structure of a liberal society (see, e.g., Rawls 1999a, 4, 98–101; 1999b, 89, 120). Re: (2), justice “is the first virtue of social institutions, as truth is of systems of thought” (Rawls 1999b, 4). Re: (3) and (4), see, e.g., Rawls’s “structuralist” response to Robert Nozick’s “transactionalism,” “The Basic Liberties and Their Priority” in Rawls (1993).

⁸ I say this because, at points, Young also seems to endorse (1)–(3) on the basis of (4). See, e.g., Young (1977, 64–74, 139, 42).

⁹ For a similar explanation, and an analysis of its role in the oppression of women, see Cudd (2006, 148–50) on the “Invisible Foot.”

¹⁰ Young (1977, 107).

¹¹ Young (1977, 196–97). I address Young’s critique of theories of discrimination in Sangiovanni (2017, ch. 3).

¹² For the idea of a participatory intention, and its connection to collective action, see Kutz (2000).

¹³ For an argument that each person does, in fact, bear *some* causal responsibility, see the helpful discussion of overdetermination in Goldman (1999). It is a further question, of course, whether we also bear moral responsibility, and why.

¹⁴ Young (1977, 100).

¹⁵ Ibid.

¹⁶ Ibid., 103–04, 7, 96.

¹⁷ Ibid., 105.

¹⁸ Parfit (1984, 79–82). I have added (a) that the switch is connected to their housing electrical circuit to make it clear that none of them acts with the intention of playing their part in torturing the man (though they foresee it); (b) changed the number of tortured individuals from 1,000 to 1 to eliminate the role of dispersion, which strikes me as irrelevant in this context; and (c) clarified that there must be at least some switches that make a perceptible difference to the man’s pain (if no one’s switch makes such a difference, then it is logically impossible for someone to be in unbearable pain at the end of the series; see Arntzenius and McCarthy 1997 for this point). Regarding (b): Parfit’s argument rests on comparing a case in which the 1,000 torturers each flip a switch 1,000 times on one of a thousand victims (a 1:1 relation) and a case in which the 1,000 torturers each flip a switch 1,000 times, increasing the pain of each victim by only a slight amount. He argues that by “pressing the button, each torturer causes each victim to suffer slightly more. The effect on each is slight. But, since each torturer adds to the suffering of a thousand victims, each torturer imposes a great total sum of suffering. Since the victims suffer just as much as they did [in the 1:1 relation], each torturer is acting just as wrongly as he used to do. In the [1:1 relation case], each torturer imposed on one victim a great sum of suffering. Each of the Harmless Torturers imposes on these thousand victims an equally great total sum of suffering” (79). But this seems implausible when there is only one torturer and 1,000 victims, the 1,000 victims are already in severe pain, and the torturer turns the dial leaving each of them imperceptibly worse off (as Parfit himself seems to concede immediately following), and doesn’t explain why it is wrong for each of 1000 torturers to flip the switch once, causing pain to a *single* victim (as in my modified example). Christopher Kutz raises the same issues with Parfit’s proposal in Kutz (2000, 130). (a) is also better because it makes it clear that the switchers have a pressing reason not to rewire, moving the example closer to the real-world cases of structural injustice we are primarily interested in.

¹⁹ See (c) in note 21.

²⁰ Young considers Kutz as the ‘best effort to extend the liability model’ but rejects his account on similar grounds as the ones offered here. She writes: “In the absence of an *intent* to produce the outcome, surely those who participate should not be found *guilty* in the same way that those who participate in a war crime are” (Young 1977, 103).

²¹ Inclusive accountability refers to the accountability we have for actions that we did not perform ourselves, but only in virtue of acting together with others. For Kutz, when someone is accountable for an action x , we can say of them not only that x is ascribable to them as a competent agent (and are therefore liable for praise or blame [whether or not such praise or blame is appropriate all things considered]), but also that they now are accountable to others for x . This usage is, for the purposes of this article, equivalent to the notion of responsibility I identify in the text.

²² Kutz (2000, 189).

²³ *Ibid.*, 186.

²⁴ Bratman (1987, Ch. 8).

²⁵ Kutz (2000, 190).

²⁶ For the first statement of the paradox, see Quinn (1990).

²⁷ See, e.g., Tenenbaum and Raffman (2005); Arntzenius and McCarthy (1997).

²⁸ Voorhoeve and Binmore (2012).

²⁹ For this point, see also Arntzenius and McCarthy (1997).

³⁰ This is better than comparing adjacent settings because you know that your ability to distinguish differences in pain is very bad when the differences are very small. More on this below.

³¹ I have calculated these figures on the basis of the assumption that the payment you require to remain indifferent between 0 and level n (given your estimates of the probability of feeling pain to a certain degree x at that level) rises at a rate of $n^{2.5}$. Note that I have disregarded the decreasing marginal utility of money to simplify things. If we assume a linear increase in utility for any given increment of money, then the “aversive” function must be convex and monotonically increasing in order for there to exist a point of “unbearability,” namely a point at which you would rather return the reward and then some to return to level 0. As I say in the text, when these assumptions are true, there must be a point of maximum expected return at some level lower than the point of unbearability but higher than 0. In our case, that is level 54.

³² They insightfully analyze the problem as depending on what Tversky (1977) calls “similarity-based” “decision-making” which can generate problematic intransitivities in pairwise comparisons. See Voorhoeve and Binmore (108–11).

³³ Consider also, in this context, that people are generally less willing to inflict pain on others than on themselves. See, e.g., the experimental evidence adduced in Crockett et al. (2014).

³⁴ If we assumed either that it is impermissible to impose any amount of discomfort for any amount of gain, then the problem is simple: everyone should rewire. But this seems implausible for very large gains and very small levels of discomfort.

³⁵ Here, I also assume that we have a moral obligation to avoid the point of “moral unbearability” rather than a moral obligation to maximize the gain. If we also believed, as some consequentialists do, that we have a moral duty to maximize the gain (assuming linear utility), then we ought to take not m , or the point of moral unbearability, into account but the point at which the cumulative gain to the whole will be maximized, which will be smaller than m . I leave this complication to the side since it does not affect the main point.

- ³⁶ This is to forestall the possibility that between levels 100 and 101, the indifference function is flat, which would imply that the probability distribution of pains felt by the man at those two levels is exactly the same. If it was exactly the same, then we could turn on our switch, but then there must be some person further down the line for whom the marginal loss would be great enough for them not to switch. (It is implausible to assume that at a certain point the pain doesn't become any worse from additional voltage, barring very extreme cases, of course, where one becomes unconscious, and so on.) If the indifference curve did eventually become flat, then, given a sufficiently large number of households, there would eventually be a point at which the aggregate gain would end up compensating for all the losses even beyond the point of unbearability.
- ³⁷ This is because the indifference curve at this point (which is beyond m) must be much steeper than the linear savings made by the household owners, so any step along that curve will represent a much larger increase than the same step on the savings line.
- ³⁸ Cf. Nefsky (2017) for an opposing view. But see the response in Barnett (2018), which supports the idea that each individual in such collective action situations *does* make a morally significant difference. See also Kagan (2011) for a different rationale that, however, also comes to the conclusion that we do make a morally significant difference in collective harm cases of the kind discussed here.
- ³⁹ Strategically, and assuming a commitment to the moral indifference function as I have characterized it, the game played by the household owners can therefore be modelled as a so-called "chicken" game in which each household owner has the following preferences in ranked order: 1. I do not rewire; fewer than m turn on their lights and so do not rewire; 2. I rewire; fewer than m turn on their lights and so do not rewire; 3. I rewire; more than m others do not rewire; 4. I do not rewire; more than m turn on their lights and so do not rewire.
- ⁴⁰ Depending on our view of moral luck, we may think that a contribution that, in fact, makes no difference to the man's pain (even though it was expected to do so) changes the character of our wrongdoing. But, even so, it would still count, on any plausible view, as an instance of recklessness, and so be an instance of wrongdoing for that reason.
- ⁴¹ Cf. Young: "I [resist]the suggestion that different people bear different degrees and kinds of [forward-looking] responsibility as contributors to structural injustice. To say that responsibility is shared means that we all bear it personally in a form that we should not try to divide and measure. That we share [forward-looking] responsibility in this way as contributors does not imply, however, that we should not distinguish degrees and kinds of responsibility in reasoning about how to take forward-looking action to discharge the responsibility" (Young 1977, 124, 37).
- ⁴² I owe this query to an anonymous reviewer.
- ⁴³ See, e.g., May (1992); Isaacs (2011) for a description of different kinds of collective responsibility *without* collective agency.

References

- Arntzenius, Frank, and David McCarthy. 1997. "Self Torture and Group Beneficence." *Erkenntnis* 47: 129–44.
- Barnett, Zach. 2018. "No Free Lunch: The Significance of Tiny Contributions." *Analysis* 78(1): 3–13.
- Bratman, Michael. 1987. *Intention, Plans, and Practical Reason*. Stanford, CA: Stanford University Press.
- Crockett, Molly J., Zeb Kurth-Nelson, Jenifer Z. Siegel, Peter Dayan, and Raymond J. Dolan. 2014. "Harm to Others Outweighs Harm to Self in Moral Decision Making." *Proceedings of the National Academy of Sciences* 111: 17320–25.

- Cudd, Ann. 2006. *Analyzing Oppression*. Oxford: Oxford University Press.
- de Sousa, Ronald. 1990. *The Rationality of Emotion*. Cambridge, MA: MIT Press.
- Goldman, Alvin I. 1999. "Why Citizens Should Vote: A Causal Responsibility Approach." *Social Philosophy and Policy* 16: 201–17.
- Haslanger, Sally. 2012. *Resisting Reality: Social Construction and Social Critique*. Oxford: Oxford University Press.
- Isaacs, Tracy. 2011. *Moral Responsibility in Collective Contexts*. Oxford: Oxford University Press.
- Kagan, Shelly. 2011. "Do I Make a Difference?" *Philosophy and Public Affairs* 39: 105–41.
- Kutz, Christopher. 2000. *Complicity*. Cambridge: Cambridge University Press.
- May, Larry. 1992. *Sharing Responsibility*. Chicago: University of Chicago Press.
- Moore Jr., Barrington. 1978. *Injustice: The Social Bases of Obedience and Revolt*. White Plains, NY: Pantheon Books.
- Nefsky, Julia. 2017. "How You Can Help, without Making a Difference." *Philosophical Studies* 174: 2743–67.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Quinn, Warren S. 1990. "The Puzzle of the Self-Torturer." *Philosophical Studies* 59: 79–90.
- Rawls, John. 1993. *Political Liberalism*. New York: Columbia University Press.
- _____. 1999a. *The Law of Peoples*. Cambridge, MA: Harvard University Press.
- _____. 1999b. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Sangiovanni, Andrea. 2017. *Humanity without Dignity: Moral Equality, Respect, and Human Rights*. Cambridge, MA: Harvard University Press.
- Sewell, William. 2005. *Logics of History: Social Theory and Social Transformation*. Chicago: University of Chicago Press.
- Tenenbaum, Sergio, and Diana Raffman. 2012. "Vague Projects and the Puzzle of the Self-Torturer." *Ethics* 123: 86–112.
- Tversky, Amos. 1977. "Features of Similarity." *Psychological Review* 84: 327–52.
- Voorhoeve, Alex, and Ken Binmore. 2006. "Transitivity, the Sorites Paradox, and Similarity-Based Decision-Making." *Erkenntnis* 64: 101–14.
- Young, Iris Marion. 2011. *Responsibility for Justice*. Oxford: Oxford University Press.