Journal of Practical Ethics

Volume 2, Number 2. December 2014

CONTENTS

The Journal of Practical Ethics is available online, free of charge, at:
http://jpe.ox.ac.uk

Editorial Policy
The *Journal of Practical Ethics* is an invitation only, blind-peer-reviewed journal. It is entirely open access online, and print copies may be ordered at cost price via a print-on-demand service. Authors and reviewers are offered an honorarium for accepted articles. The journal aims to bring the best in academic moral and political philosophy, applied to practical matters, to a broader student or interested public audience. It seeks to promote informed, rational debate, and is not tied to any one particular viewpoint. The journal will present a range of views and conclusions within the analytic philosophy tradition. It is funded through the generous support of the *Uehiro Foundation in Ethics and Education*.

Copyright

ISSN: 2051-655X

How Theories of Well-Being Can Help Us Help

VALERIE TIBERIUS

University of Minnesota

ABSTRACT

Some theories of well-being in philosophy and in psychology define people's well-being in psychological terms. According to these theories, living well is getting what you want, feeling satisfied, experiencing pleasure, or the like. Other theories take well-being to be something that is not defined by our psychology; for example, they define well-being in terms of objective values or the perfection of our human nature. These two approaches present us with a trade-off: The more we define well-being in terms of people's psychology, the less ideal it seems and the less it looks like something of real value that could be an important aim of human life. On the other hand, the more we define well-being in terms of objective features of the world that do not have to do with people's psychological states, the less it looks like something that each of us has a reason to promote. In this paper I argue that we can take a middle path between these two approaches if we hold that well-being is an ideal but an ideal that is rooted in our psychology. The middle path that I propose is one that puts what people value at the center of the theory of well-being. In the second half of the paper I consider how the value-based theory I describe should be applied to real life situations.

INTRODUCTION

Well-being is, by definition, what is good for you. If you achieve well-being in your life, you may not have lived a morally perfect life and your life may not have

made any great contribution to art, world peace or progress, but you will have lived a life that is good for you. Even though a good life in this sense is not the same as a perfect life (whatever that might be), well-being is still an ideal. It is something we strive for and we certainly do not all achieve it. Our well-being may be diminished by health problems, bad financial luck, the death of a loved one, poor planning, or many other factors. Even if we are lucky and things go well for us, the ideal of a good life serves as a goal for our aspirations about how things might go even better.[1]

There are a variety of different theories of well-being in philosophy and in psychology that take well-being to be an ideal to different degrees. Some theories define well-being in terms of people's psychology to a much greater degree than others. Theories that define well-being in terms of our psychology directly keep the ideal down to earth. Other theories define well-being in terms of objective values or the perfection of our human nature and these theories let the ideal move farther away from people's actual psychological perspective. These two approaches present us with a trade-off: The more we define well-being in terms of people's subjective psychological states, the less ideal it seems and the less it looks like something of value that could be an important aim of human life. On the other hand, the more we define a person's well-being in terms of objective features of the world that do not have to do with his or her psychological states, the less it looks like something with which a person should obviously be concerned or something he or she has a reason to promote.

What I want to argue in this paper is that we can take a middle path between these two approaches if we say that well-being is an ideal—something it makes sense to say is valuable—but an ideal that is anchored in our psychology. Other theories have taken this path. Full information theory, for instance, defines well-being in terms of idealized psychological states, namely the desires that we would have if we were fully informed. I believe such theories are on the right track, but I also think that existing theories of this kind can be improved upon. In this paper I propose a version of these idealized subjective theories that I hope shares their virtues and avoids their shortcomings.

Idealized subjective theories in general have the problem that we do not have ideal psychologies to work with, which means that there are special difficulties for

1. At least by my definition of well-being. There is some controversy about how this and related concepts should be used, but these controversies need not concern us here. The good for a person is what I mean to be talking about and the particular word used to refer to it is not of great importance.

applying such theories of well-being. If we don't have access to what our psychological states would be like *ideally*, how do we help promote well-being defined in terms of such states? I answer this question by articulating a different way that a theory of well-being can be helpful. Instead of providing us with a detailed picture of all the elements of an ideally good life, I argue, a theory of well-being can give us practical guidance about how to change a person's life so that it improves. In other words, a theory of well-being can fulfill its practical function by instructing us about the process of improving people's lives rather than by giving us a sharp picture of the ultimate goal.

So, this paper has two aims: first, to describe a theory of well-being that strikes the right balance between real and ideal, and second, to show how this theory can be applied to the practical matter of helping improve people's well-being. In the first section of the paper I will explain in a little more detail the background that I've gone over quickly in this introduction. In section two, I will outline the theory I favor: the value fulfillment theory of well-being. In section three I discuss how the theory can be applied.

REAL OR IDEAL?

Some theories define well-being in terms of our actual psychological states. Many psychologists, for example, think that well-being consists in life satisfaction and positive affect balance (roughly, more pleasant feelings than painful ones) (Diener 1984; Diener 2006). Some philosophers agree that well-being should be defined in terms of mental states like pleasure and pain. According to hedonism about well-being, the good life for a person is a life that has the most pleasure and the least pain (Crisp 2006; Feldman 2004). Others (many philosophers and economists) think that desires or preferences are the right psychological state to focus on. According to the desire satisfaction theory of well-being, the good life for a person consists in getting the most of what she ultimately wants over the course of her lifetime (Heathwood 2006; Heathwood 2005).

All of these theories have something going for them and it is not the purpose of this paper to show that these theories are wrong. What I want to point out is that these theories make well-being depend very heavily on our individual psychologies. What we happen to take pleasure in, to be satisfied by, or to want fundamentally determines what is good for us. It's not that these theories don't give us any ideal

to which to aspire at all, but rather that the ideals these theories posit are defined in terms of each person's actual psychology. For instance, the ideally good life according to desire satisfaction theory is not the life that many people actually achieve (few of us are able to get all the things that we want), but it is an ideal that is fixed by what we really do want.

Though these theories do give us something of an ideal, many will find these ideals wanting. Well-being is supposed to be one of the main goals of human life, that at which we aim in deliberation and planning when we think about how to live our lives. Could the mere satisfaction of our desires play such a role? Think about someone whose desires seem ill suited to living a good life, for example, someone who desires nothing but money and power, or a person with anorexia nervosa who desire to be thin above all else. We might think that a theory of well-being ought to allow us to question whether satisfying these desires really is good for a person in any way, but actual desire satisfaction theory does not allow this.[2] Or think about the well-being of children. People tend to think that part of what it is to raise children well is to instill the right desires in them so that they want to be productive, decent people. If well-being is just desire satisfaction, it is unclear where these standards for the "right desires" will come from.

Other theories of well-being allow the ideally good life for a person to move farther away from her psychology. According to eudaimonism, for instance, the good life for a person is the one in which she fulfills her human nature, where what counts as a person's nature has much to do with what is normal for members of the human species not with what this particular person happens to like (Foot 2001). Objective list theories of well-being say that a good life for a person is one in which she achieves certain objective goods such as friendship, knowledge and pleasure (Arneson 1999; Finnis 1980). Such theories make well-being an ideal that could be far removed from what a person actually thinks about what is good for her. If she is different from other human beings, or doesn't care about certain objective values, for example, the way that the theory defines well-being for her might not be something she has any real interest in pursuing.

We can now see more clearly the trade-off that I mentioned in the introduction. Theories that make well-being a function of our actual psychology do not

2. Actual desire theory does allow some room for criticizing defective desires, for instance, on the grounds that satisfying one will cause one to have less overall desire satisfaction in the long term. See Heathwood (2005).

explain why well-being is a valuable goal of human life. Theories that idealize well-being away from our actual psychology do not explain why well-being should be *our* goal.

THE VALUE FULFILLMENT THEORY OF WELL-BEING

There are surely many ways of resolving this problem. To argue that one way is better than any other possible way is far beyond the scope of a single paper. Instead, I will take an approach that has seemed promising to many and put a new spin on it that makes it an even more compelling solution. In doing so, my starting assumption is that a theory of well-being must explain why well-being is a valuable ideal and also why it is a valuable ideal for each of us.

The promising approach I have in mind defines well-being in terms of a person's *ideal* psychology: for example, theories according to which well-being consists in getting what you would desire if you were fully informed and rational, or what your fully informed self would want you to want, or what you take to be a satisfying life insofar as your assessment is fully authentic (Railton 1986; Griffin 1986; Brandt 1979; Sumner 1996). Such theories promise to explain how well-being is specially related to individual subjects, because they appeal to an individual's desires or satisfactions. They promise to explain how well-being is something valuable, because they do not take our desires and satisfactions at face value, but rather as these desires and satisfactions might be improved in accordance with norms of improvement (such as rationality or authenticity).

It seems to me that the promise of these idealized subjective theories (as I called them in the introduction) has not been fully appreciated. One reason for this is the serious objections to full information as a norm of improvement. Philosophers have argued that the ideal is at best alienating and at worst incoherent (Rosati 1995; Velleman 1988; Tiberius 1997). Another reason has to do with the psychological states that have been at the center of these theories; critics have argued strenuously against the relevance of desire and life satisfaction to well-being (Richard Kraut 1994; Haybron 2011). The theory I propose is an idealized subjective theory that takes values (rather than desires or satisfactions) as the key psychological state, and a model of a value full life (rather than an informed or authentic agent) as its ideal. In the remainder of this

section I will explain the theory in more detail, in the hope that a good description of it will reveal its advantages.[3]

Let's think first about which psychological states a theory of well-being should concern itself with? Preferences? Pleasures? Life satisfaction? I believe that the aspect of our psychology it makes most sense to attend to in our theories of well-being is our *values*. This is because values are what people themselves take to be relevant to how their lives are going; our values are the goals we plan around and use to assess how well we're doing in life. For this reason, a theory of well-being that focuses on what people value is well suited to explain why well-being is something that people have a particular reason to care about. Moreover, values are held to standards in ways that desires or pleasures are not; it makes sense to talk about what it is appropriate to value and we tend to think that we should have reasons for valuing what we value. This gives values a leg up when it comes to well-being, because it allows them to make sense of how we can go wrong in pursuing our well-being. Accounting for how we could go wrong or make mistakes about what is good for us is needed to make sense of well-being as a normative notion.[4]

To value something is, in part, to be motivated with respect to it; desires and values are similar in this respect. But values have a special status in our planning and evaluation, they have greater stability than mere preferences and they are emotionally entrenched in ways that desires might not be. For example, a person who values being a parent will be disposed to make plans that include spending time with her child, to feel joyful when she spends time with her child, disappointed when she misses her child's ballet recital, and so on. She will also be inclined to take into account how well she is doing as a parent when she thinks about how well her life is going and how she could improve. In short, then, values are what we value, and to value is to have a co-ordinated pattern of emotions and motivations toward something that you take to be relevant to how your life goes. Not all values are fully realized—sometimes our motivations to act, our emotions and our judgments are out of sync with each other —but values in their most complete sense include all these elements. Values, as I intend them, then, are relative to subjects; different people may value different things. That

3. I defend a close relative of this view in more detail in Tiberius 2008.

4. This is an important topic in its own right and more needs to be said about how the value fulfillment theory makes sense of the possibility of error. I will say a little more about this shortly, but my main focus in this paper is on how the theory can be used for guidance.

said, there are many shared values, especially when it comes to relatively basic values: almost everyone values health, happiness, friendship, family and meaningful work.[5]

Now that we know what values are we can see how they fit into what I call the value fulfillment theory of well-being or VFT. According to VFT, a person's life goes well to the extent that she pursues and fulfills or realizes things that she values where those values are emotionally suitable, mutually realizable and seen by the person to make her life go well.[6] The best life for a person is the one in which she gets the most value fulfillment she can, given her personality and environment, and what is good for you now is to do what contributes to some specification of the best, "value full" life. In short, we live well when we realize what matters to us over time. This includes achieving certain states of affairs (such as career goals) and also maintaining the positive affective orientation that comprises valuing something. If your (suitable and realizable) values include your own enjoyment, relationships with family and friends, accomplishing something in your career, and contributing to certain morally worthwhile projects, then your life goes well for you insofar as you have good relationships and career success, make a moral contribution and enjoy what you're doing, as these continue to be the things you care about.

What it is for a value to be fulfilled or realized and what it means to say that one life has more value fulfillment than another are obviously very important for VFT. Values, like desires, bring with them standards for success, and living up to these standards is part of value fulfillment. These standards are not always as obvious; some values are such that we succeed in their terms by having the right attitudes or being a certain kind of person. Nevertheless, there are standards for values in the sense that there are ways of responding appropriately or inappropriately given the nature of what is valued (see Anderson 1995). Moreover, most values encompass standards that are objective in the sense that whether or not we fulfill them is not a matter of whether we believe we are fulfilling them. There is something to meeting the standards that our values impose that goes beyond our subjective experience. In this respect, value fulfillment is similar to preference satisfaction: you may fail to get what you want without knowing it (say, if you are seriously deluded), and you may fail to fulfill your values, though you believe otherwise. Finally, if we are going to achieve what matters to us, it is not only success in terms of what is valued that matters, but

5. For a more thorough discussion of this view of values and the research on what people value see Tiberius 2008.

6. For a similar approach to the relationship between values and well-being see Raibley 2010.

also the valuing attitudes themselves. We require some stability in our valuing attitudes if we are going to succeed by the standards we think are important. (Of course, there is such a thing as too much stability: how much stability is required, and when change is recommended, are difficult practical questions, as we will see in the next section). Value fulfillment, then, is succeeding by the standards of your values while continuing to think that these standards are important to how well your life goes.

Assessing *total* value fulfillment requires attending to the relationships between values. People's values are typically complex. We value some things largely as a means to others (for example, you might value running marathons as a means to the values of health and fitness). We value some things as constitutive of other more abstract things (for example, you might value playing the piano *as a way of* valuing music). Some values are more important to us than others and some values have a more central role in the whole system. These considerations must be taken into account when we evaluate total value fulfillment and we ask whether one life has more overall value fulfillment than another. Importantly, it is not necessarily the case that getting more fulfillment of a single value at the expense of fulfilling others to a smaller degree contributes to the best overall life. This is because of the ways in which values are related to each other. Consider a simple example to illustrate this point. Imagine Bob, a person whose main values are meaningful work and family life. As with most people, Bob finds that these two values often conflict with each other because of the amount of time they each demand. You might think that VFT implies that Bob would be better off quitting his job and attending to his family, or leaving his family and focusing on his career, but VFT implies no such thing. First of all, if work and family are really both important to Bob, he might very well get more total fulfillment by achieving each of these values to a lesser degree than he would by achieving either on its own. But more importantly, for a normal human being like Bob it is very unlikely that he could make great strides in one if the other were entirely abandoned. This is partly because of diminishing returns (working all the time often does not lead to progress). And it is partly because of the role of other values that Bob (like most normal human beings) has: Bob's health would likely be affected by his working all the time and not developing close personal relationships, his enjoyment would likely be decreased by spending all his time in one way, and so on.

We can now see how the value fulfillment theory promises to accommodate both sides of the trade-off for theories of well-being. It defines well-being in terms of a person's individual psychology, namely, her values. But it also posits an ideal

and allows for the possibility that a person's psychological states are in need of improvement or transformation (thus allowing for the possibility of error). For example, the person with anorexia nervosa has values that are just not conducive to a value full life, since the value of thinness competes with other values (physical and mental health) and even with life itself (a necessary pre-condition for value fulfillment). The compelling ideal of a value full life—a life in which we do well by what matters to us—does constrain which values it makes sense for a person to have. Nevertheless, the ideal does not impose external values on a person in a way that risks its appearing unrecognizable to someone as what is good for him or her.

APPLYING THE VALUE FULFILLMENT THEORY: FROM IDEAL TO REAL

One problem with idealized subjective theories is that we do not have access to our ideal psychological states and this makes it difficult to apply such theories to real life problems. The value fulfillment theory is certainly not immune to this difficulty; indeed, in some ways the focus on values and the ideal of a value full life makes the problem worse. The ideal of a value full life provides guidance for thinking about what a good life is through the standard of the fulfillment of a set of values over time, but even with these guidelines about what counts as fulfillment, there are many different ways of living a life in which you value and have good friendships, meaningful work, enjoyable experiences, and so on. The complexity of systems of values and the fact that values themselves are open to interpretation mean that there will be no single, well-defined best life for a person overall or even at a particular time. This is in part because the "units" of value fulfillment are large and in part because there are different ways that values can be successfully organized even for a single person. If the units of fulfillment were small, we could rank possible lives in terms of minute gains and losses. If there were only one way for a particular set of values to be realized together, then there would be a clear sense in which there is a best life for a person. But this is not how values are. Instead, the value fulfillment approach tells us that the good life for a person overall is one of the lives in a set of roughly equivalently value full lives that constitutes a model of a good life for a person.[7]

7. Raibley (2012) uses the notion of a "paradigm" where I prefer to talk about a model. I think it is just as useful and perhaps a bit more precise to think about a set of best lives for a person that is a model of well-being.

There are, in other words, many different shapes that the ideal of a value full life can take and, to make matters worse, what is in that set of value full lives will change over time as the person makes choices that close off some options and open others.[8] It's easy to see this when it comes to career choices. There is a point in one's life when the value of meaningful work could be specified in many different ways, constituted by many different kinds of work. But as a person ages, acquires training and specialization, the options for living a life with the most value fulfillment change. Whereas there is a time at which being a teacher, doctor or baker could all have roughly equal value fulfillment, once you have spent 20 years practicing law the calculation is not the same. This is certainly not to say that making large changes can never improve your life, however, the amount of value fulfillment you can expect by quitting your job as a lawyer mid-career and going to medical school is different from the amount of value fulfillment you can expect as a young person deciding between medicine and the law. Similarly, as anyone who has children will tell you, once you have children your values change profoundly; you suddenly value your child, your relationship with him or her, and your identity as a parent. Therefore, once you have children, it is almost certainly true that all of the lives that have the most value fulfillment for you are lives in which your children are healthy and happy and you enjoy being a parent to them. But for many people, before they have children there are value full lives open to them that do not include having children.

Applying the theory, then, is not going to be a simple matter. But we can make progress by thinking of the practical contexts in which such applications take place. What practical purposes do theories of well-being have? For what purpose would we need to translate the ideal life given by a theory of well-being into reality? Basically, we need to bring the ideal down to reality when we want to help somebody (or help ourselves), to make their (or our) lives better. It is as potential benefactors that we try to discern the exact shape of a good life, and this endeavor takes place in a particular context that determines whom we aim to help and in what way. (In what follows I focus mainly on friends as potential benefactors and beneficiaries, but the points I'll make can be extended to other relationships. Benefactor and beneficiary could even be the same person if the context is of someone who is trying to evaluate and improve her own life).

According to the value fulfillment theory, there are some broad guidelines

8. Of course, this is true for desire satisfaction theories too, but it is not often noticed as a challenge for the application of theory.

for what a benefactor should attend to in almost any context: if you are trying to assess how well a person is doing you will need to ascertain (a) what that person's core values are[9], (2) how well she is succeeding in terms of these values, and (3) how likely it is that the status quo will lead to a life of high total value fulfillment over time. To figure out how the person's life could be improved you also need to think about how things could be different such that greater total value fulfillment will be achieved. This requires thinking not only about what kinds of core values and specifications of cores values would be more compatible and more likely to lead to fulfillment, but also about what kinds of changes the person is actually capable of making and how she sees her own good. For example, consider Jane, a person who values creativity and accomplishment, but who has manifested these values in her life with a career for which she has no talent. Let's say Jane has dedicated herself to writing novels, but she is destined to experience only frustration as a writer. To help Jane we need to think about how else these values can be realized in a life and whether she could become a person who fulfills these values in a different way, say through practicing an art for which she has more talent, or by seeing the creative aspect of something for which she does have talent.

These things are not easy to figure out, but they do seem like the right things to think about when we aim to understand how someone is doing and to help improve her situation: how is she doing with respect to what she cares about, does she care about the right things given her personality and circumstances, and could her situation or her values be changed so as to make her life one in which she is better able to achieve what matters to her. When we see that such assessments of how people are doing take place in a practical context, new challenges arise. These challenges give rise to more guidelines for the process of translating the ideal into reality.

The first kind of challenge is epistemological: there are a variety of things that the benefactor might not know that will affect his or her ability to assess how much the beneficiary's life resembles the ideal and how it could be improved. In particular, the benefactor needs to know (and might be wrong about): what a good life is for the person in broad terms, how the ideal could be specified, what changes are required for bringing about the better life, and what changes the person is capable of making.

9. "Core values" are values that are more important, more central and/or more likely to be at least in part valued intrinsically. I think these features of values are a matter of degree; so, a particular value can be more or less core. Core values are often very general and need to be instantiated or specified in some particular way to be pursued

No one (including the person herself) could possibly have perfect knowledge about all of these things, but some will have better information than others.

There is no easy solution to the epistemological problem. The basic guidelines for potential benefactors are to try to acquire more knowledge, to proceed on the basis of conservative assumptions about what almost everyone values, and to proceed with caution given the possibility that we do not know how to help. On the first point, we can think about acquiring knowledge about the particular person or group of people we are trying to help. It is also helpful to learn about human psychology in general: for instance, how human beings tend to reveal what they really value (what clues to look for) and how people are able (or unable) to change course in life. Learning about human nature generally can help us make reasonable assumptions about what core values individual people are likely to have. (Note, however, that it is still the individual person's relationship to the value—not the human species' relationship to it, as eudaimonism would have it—that explains why it is good for her). Many of the values people have are socially sanctioned, highly stable and abstract enough that how they are fulfilled is open to interpretation: for example, health, pleasure, close family ties, friendship, comfort and security. These values are quite likely to be a part of any of the best lives a person could live (though particular means to them might differ) and therefore they form an excellent basis for well-being assessments. We can also assume that values that are related to other values in fundamental ways—as necessary conditions for their pursuit (such as health), or as a justification for other values (such as psychological happiness)—will be stable and emotionally appropriate over the long term. A useful heuristic, then, in assessing well-being and making judgments about how to benefit people, is to pay attention to the basic values that are likely to be part of the best life for anyone. We shouldn't underestimate how far this takes us.

The fact that we might lack crucial information about a person's situation or the ideal life for her is one significant problem, but there is a second kind of challenge that is independent of our epistemic position. Even if you are correct about what is good for a person and what is wrong with the way she is currently living her life, intervening in someone's life on behalf of values that you think would be more appropriate for them introduces costs in value fulfillment terms: ruptures to the bonds of friendship, pain and dissatisfaction. Most of us have had friends who pursue romantic relationships that are bad for them, but it is rarely a good idea to take drastic measures to prevent our friends from making dating mistakes. A friend who

criticizes our choices too much is not one we are likely to confide in or turn to for help when things turn out badly. Overriding or ignoring the actual values of someone with whom one has a professional (rather than friendly) relationship in order to help that person causes its own problems such as the feeling of being disrespected, the erosion of trust in helping professionals, and the violation of role defined obligations. Finally, success by the standards of an inappropriate value is not always entirely bad for a person. This is because of the relationship between values: succeeding in any terms (even if the value achieved is not perfectly appropriate for you in the long term) usually brings pleasure, a sense of satisfaction or accomplishment, and other valued rewards. For all these reasons, the fact that a person's values could be better for her does not license us to ignore her actual values in the usual case.

It does seem, though, that there are some contexts in which it makes sense to discount a person's inappropriate values in practice, such as when these values are so dysfunctional that they do not have any connections to other values and rewards, or when the risks of negative consequences are minimal. For example, a friend who is putting herself and her children at risk by staying in an abusive relationship almost certainly needs to adjust her values (on the assumption that she does value the relationship, which of course she may not). In this case, continuing to value the relationship with the abusive partner is at odds with other things she values more, or should value more if she is going to live a life of high value fulfillment. It might indeed be a duty of friendship to try to intervene in some way or another with this friend's situation. To take a less dire example, a student who asks for guidance from a teacher might be well served by advice that recommends changing her goals.

The second type of challenge, then, is the challenge of ascertaining whether it is desirable (in terms of the goal of promoting well-being) to discount, ignore or override a person's actual current values. Let's call this the interpersonal challenge. The interpersonal challenge also does not admit of an easy solution. But we can minimize the risk of making things worse by taking care to assess the circumstances before trying to help. Some circumstances make it more appropriate to discount, ignore or override a person's values than others. One factor that is relevant to any decision under uncertainty is the degree of risk of harm or benefit. If the beneficiary's values are a clear and present danger to her, it makes more sense to ignore these values than if her values are just somewhat less than ideal. For example, in order to benefit a person who has become addicted to drugs or has fallen into a cult (long enough to say that she genuinely values participation in the cult) it might be necessary to ignore

or override her current values because they put her health and even her life in danger. Second, it matters whether the beneficiary could actually change so as to have better, more ideal, values. If Jane is just never going to give up on her dream of being a writer, it might do no good for you to try to help her by engaging in a long term project of talking her into doing something else.

A third important factor is the kind of relationship that exists between the benefactor and the beneficiary. What kinds of relationships make this sort of helping appropriate? I suggest that there are several important variables. Trust (that the person trying to help you has your best interests at heart) is important because it makes some room for honesty about what might be wrong with a person's life and it makes it more likely that the beneficiary will accept (and hence benefit from) the benefactor's help. A certain level of intimacy or understanding is also required among friends without which advice about how to live one's life better might be taken to be nosy or condescending. Trust and intimacy make a difference to whether someone will be benefitted by help that assumes the need for a transformation of actual values. Making changes to our values is difficult, so we need to have some reasonable prospect that a proposed change would be good for us before we'll put in the effort. Advice from a person who knows us well (intimacy) and whom we trust to care about how well our lives are going is, prima facie, better than information from someone who knows us less well or whom we suspect of having ulterior motives. Another important variable is the extent to which the friends' lives are intertwined. One thing that makes it more acceptable for a life partner to criticize the spouse's actual values is that the two of them have to live together and they share each other's burdens to a greater extent than most friends. A final important factor is the skill that the friend has in communicating difficult or sensitive information. Some ways of telling a person that her goals are inappropriate and likely to make her life go poorly over the long term are more tolerable than others, and a more effective benefactor has the skills to communicate this information in the best way.

No one of these variables is sufficient to determine whether it's appropriate to discount a beneficiary's values and there's no algorithm for how much of each variable is needed. You might have an intimate relationship with someone in the sense that you have a long history together and know each other to the core, but without trust such a friend's suggestions about how we could better our lives are not helpful. Think of divorced partners who might know each other better than anyone else in the world, but who are in no position to give advice because there is no longer a

presumption that each has the other's best interest at heart. Trust without intimacy is also not sufficient, because a trusted person who doesn't know you very well can't make good assessments about how you would be better off. For example, many of us have parents who value our welfare more than anything else, but are not well positioned to help us live better lives because they still see us as the children we once were. Trust and intimacy without skill are also problematic: someone who wants to help and knows enough about you to have a useful perspective, but who doesn't have the sensitivity to communicate that perspective in a way you can accept isn't in the best position to help you overcome your dysfunctional values. Notice that the presence of certain formal relationships, such as the relationship between patient and therapist, can change the required balance of qualities: a therapist who is trusted and skilled can do with less intimate knowledge of the patient in part because he or she will have greater general knowledge of human psychology and professional knowledge about the patient's life. Finally, mutual dependence weighs against shortcomings with respect to the other factors, because the costs of sticking with the status quo might be very high, but mutual dependence by itself is no guarantee if the other factors are absent.

We have discussed a number of conditions for the appropriateness of discounting, ignoring or overriding a beneficiary's current values for the sake of helping to promote her well-being. To summarize, it can make sense for a benefactor to discount, ignore or override a beneficiary's actual values in assessing or trying to improve her well-being under four conditions:

1. The beneficiary's values are truly harmful

2. The beneficiary could change

3. There is an appropriate relationship between the helper and the beneficiary, defined in terms of: intimacy, trust, skills of communication, and the extent to which lives are intertwined

4. The helper is in a good epistemic position with respect to the above.

These conditions are, in essence, guidelines for applying an ideal theory to a real life in practice. More guidance, as we have seen, is found in the value fulfillment the-

ory's account of what counts as a value full life. It is worth pointing out that there are many different ways of trying to help that might discount, override or ignore a beneficiary's current values. One can give advice, provide alternative options, withdraw support, directly intervene, coerce or force. These different actions vary in terms of how intrusive they are and the more intrusive, the higher the stakes. As one contemplates more intrusive actions for the sake of a person's welfare, the above conditions become more stringent.

Certainly, the story the value fulfillment theory tells about how to benefit people is not a simple one. If this theory is correct, it turns out that the practical application of the theory of well-being requires those who want to help other people to figure out how those people's lives could be closer to an ideal, where there are many different shapes the ideal could take in practice. Is this a problem with the theory? I don't think so, for two reasons. First, according to the value fulfillment theory (and, indeed, any plausible theory of well-being) there are many very easy ways to help very many people. Almost everyone values health and enjoyment for themselves and their friends and family. Deprivations that make it impossible to attain these values are an obvious road block to achieving well-being. In particular, people who are suffering from illness, who are in pain, or who lack basic material resources (of whom there are vast numbers in the world) can be helped tremendously by alleviating these impediments to living value full lives. There are even easy ways to help people who are more fortunate in terms of their basic needs, because there are many cases in which helping people to pursue their valued projects is exactly the best way to help improve their lives.

Second, for the other cases—cases in which a person is not deprived of the basic necessities that make it possible to live in accordance with her values and has values that are harmful to her in some way—it should not be surprising that it is not easy to help. To see why not, think of a paradigm type of case of dysfunctional values, a case in which core values conflict. For example, consider Joe, the gay evangelical Christian. Joe deeply values his religious identity and his church, and yet his sexual identity is completely rejected by this church. If Joe also values having satisfying romantic relationships, he is in trouble with this set of values. If you are Joe's friend, how should you help him? Given my own beliefs about religions like this, if I were Joe's friend I would be likely to try to talk him into joining a different church. But is it completely obvious that this is the best way to help Joe? What if he is unable to get rid of the belief that living as a gay man would result in his eternal damnation? What if

he does not actually care that much about romantic relationships? What if by telling him to join a new church I reveal that I have no understanding of his religious identity at all and he loses confidence in one of the few people in his life with whom he can share this problem? My point here is not that one shouldn't make some effort to get Joe to change his mind about his particular church (or about eternal damnation). Rather, my point is that it's hard to know what to do for Joe.[10] Cases like these are difficult and so it is no criticism of the value fulfillment theory that it acknowledges this fact of life. Indeed it is a point in the theory's favor that it helps explain why such cases are so difficult.

CONCLUSION

The value fulfillment theory says that to live well is to succeed in terms of our own values. The best life we can live (in terms of our own well-being) is the one in which we get the most value fulfillment overall, and what is good for us to do now is whatever contributes to living a life that is closer to this ideal, which will sometimes require changing our values in some way. The value fulfillment theory does make well-being an ideal, though the ideal is relative to the evaluative outlook of the person. Therefore, well-being is both ideal and psychological. Unlike other theories that define well-being in terms of our idealized psychological states, VFT does not propose particular norms for the improvement of individual psychological states such as full information, rationality, or authenticity. Rather, it asks that we evaluate our current values by comparing them to an ideal of life in which we succeed in terms of the standards imposed by what we care about over the long term. Since there are many paths to an ideal life for a person that change as life goes on, applying this theory to real life is a challenging process. A theory of well-being can help us in this process by identifying the standards of success that we should employ, and the dangers we should aim to avoid, when we are assessing how well people are doing and imagining how they might do better.

It is impossible to provide a complete defense of a theory of well-being in a single paper and I have not tried to do that. I hope to have highlighted some of the advantages of thinking of well-being in terms of values and ideals of value fulfillment,

10. I was prompted to think about the difficulties involved in this kind of case by an article in the New York Times Magazine entitled "Living the Good Lie" (Swartz 2011).

however, and to have addressed one of the main concerns that arise for the application of a theory like this.

REFERENCES

Anderson, E. 1995. *Value in Ethics and Economics*. Harvard University Press.

Arneson, R. J. 1999. "Human Flourishing versus Desire Satisfaction." *Social Philosophy and Policy* 16: 113–42.

Brandt, R. B. 1979. *A Theory of the Good and the Right*. Oxford University Press.

Crisp, R. 2006. "Hedonism Reconsidered." *Philosophy and Phenomenological Research* 73 (3): 619–45.

Diener, E. 1984. "Subjective Well-Being." *Psychological Bulletin* 95 (3): 542–75.

———. 2006. "Guidelines for National Indicators of Subjective Well-Being and Ill-Being." *Applied Research in Quality of Life* 1 (2): 151–57.

Feldman, F. 2004. *Pleasure and the Good Life.* Clarendon Press.

Finnis, J. 1980. *Natural Law and Human Rights*. Clarendon Press.

Foot, P. 2001. *Natural Goodness*. Oxford University Press, USA.

Griffin, J. 1986. "Well-Being: Its Meaning." *Measurement, and Moral Importance.* Clarendon Press.

Haybron, D.M. 2011. "Taking the Satisfaction (and the Life) out of Life Satisfaction." *Philosophical Explorations* 14 (3): 249–62.

Heathwood, C. 2005. "The Problem of Defective Desires." *Australasian Journal of Philosophy* 83 (4): 487–504.

———. 2006. "Desire Satisfactionism and Hedonism." *Philosophical Studies* 128 (3): 539–63.

Kraut, R. 1994. "Desire and the Human Good." In *Proceedings and Addresses of the American Philosophical Association*, 68:39–54.

———. 2009. *What Is Good and Why: The Ethics of Well-Being*. Harvard University Press.

Raibley, J. 2010. "Well-Being and the Priority of Values." *Social Theory and Practice* 36 (4): 593–620.

Railton, P. 1986. "Moral Realism." *The Philosophical Review* 95 (2): 163–207.

Rosati, C. S. 1995. "Persons, Perspectives, and Full Information Accounts of the Good." *Ethics* 105 (2): 296–325.

Sumner, L. 1996. *Welfare, Happiness, and Ethics*, Clarendon Press.

Swartz, M. 2011. "Living the Good Lie." *The New York Times*, June 16, sec. Magazine. http://www.nytimes.com/2011/06/19/magazine/therapists-who-help-people-stay-in-the-closet.html.

Tiberius, V. 1997. "Full Information and Ideal Deliberation." *The Journal of Value Inquiry* 31 (3): 329–38.

———. 2008. T*he Reflective Life: Living Wisely with Our Limits*. Oxford University Press, USA.

Velleman, J. D. 1988. "Brandt's Definition of 'Good.'" *The Philosophical Review* 97 (3): 353–71.

What Can We Learn From Happiness Surveys?

EDWARD SKIDELSKY

University of Exeter

ABSTRACT

Defenders of happiness surveys often claim that individuals are infallible judges of their own happiness. I argue that this claim is untrue. Happiness, like other emotions, has three features that make it vulnerable to introspective error: it is dispositional, it is intentional, and it is publically manifest. Other defenders of the survey method claim, more modestly, that individuals are in general reliable judges of their own happiness. I argue that this is probably true, but that it limits what happiness surveys might tell us, for the very claim that people are reliable judges of their own happiness implies that we already have a measure of how happy they are, independent of self-reports. Happiness surveys may help us extend and refine this prior measure, but they cannot, on pain of unintelligibility, supplant it altogether.

Measurements of self-reported happiness are taken increasingly seriously by psychologists, sociologists and (more recently) by economists. They form part of the official statistics of many nations. Yet they remain beset by methodological problems. Some of these are superficial. For instance, people's satisfaction with their life as a whole can be significantly influenced by trivial recent events, such as finding a dime (See Schwarz and Strack, 1999, p. 62). This kind of "noise" can be eliminated by good survey design. Other problems are more intractable. It has been said, for instance, that studies claiming to show that English people are happier than Poles reveal nothing more than the fact that the English word "happy" is used more freely and lightly than its Polish counterpart (See Wierzbicka, 2004). Disentangling such semantic effects

from real differences of happiness is on-going problem for compilers of international happiness statistics.

Interesting and important though these quandaries are, I am here going to set them aside in favour of a philosophically more basic question: do people *know* how happy they are? If the answer to this question is negative, the whole project of measuring happiness by means of self-reports is in jeopardy.

"Do people know how happy they are?" could mean one of two things. It could mean, "Are people infallible judges of their own happiness?" Or it could mean, "Are people in general reliable judges of their own happiness?" Taken in the first sense, I argue that the answer to the question is "no". Taken in the second sense, I argue that it is probably "yes". However, this is not the unequivocal vindication of the survey method that it might at first appear, for the very claim that people are reliable judges of their own happiness implies that we *already have* a measure of how happy they are, independent of what they tell us. Happiness surveys may help us extend and refine this prior measure, but they cannot, on pain of unintelligibility, supplant it altogether.

ARE PEOPLE AUTHORITATIVE JUDGES OF THEIR OWN HAPPINESS?

It is often asserted that individuals are the ultimate arbiters of whether or not they are happy. "If people say they are happy then they *are* happy," writes Michael Argyle in a foundational textbook on the psychology of happiness. "If people say they are depressed then they *are* depressed." (Argyle, 1987, p. 2) In a similar vein, psychologist David G. Myers has written, "by definition, the final judge of someone's subjective well-being is whomever lives inside that person's skin. 'If you feel happy,' noted Jonathan Freedman ... 'you are happy – that's all we mean by the term'"(Myers, 2000, p. 57).

Not all psychologists are so confident. Daniel Kahneman has famously claimed that I can be mistaken about my overall happiness level. But this is only because he believes that my overall happiness in a period is a function of my happiness at each moment in that period, and that I can fail to recall this latter accurately. His doubt, in other words, concerns memory. He does not think that I can be mistaken about how happy I am *right now*. Hence Kahneman has suggested that we can get a more accurate

measure of people's overall happiness level by asking them how happy they are hour-by-hour over a several week period and integrating the results (Kahnemann, 1999).

Whence this trust in first-person happiness reports? The answer, I suspect, is a certain picture of the mind familiar to philosophers from Descartes and many others following him. Mental states, on this picture, are indubitable, meaning that if one experiences them one cannot doubt that one experiences them. The mind is transparent to itself. This picture is a thoroughly misleading one. Indubitability is characteristic of some, but not all mental states. One cannot doubt that one is in pain or seeing red. But one can, notoriously, doubt whether one is in love or believes in God. One can also, I shall argue, doubt (and be mistaken about) whether or not one is happy.

In a sense, we all know this already. It is a point of common knowledge that many people, particularly young people, fancy themselves to be deeply unhappy when they are not really unhappy at all. The difficulty is making sense of this commonplace thought. What kind of thing must happiness be if it is possible to be mistaken about whether one is happy?

A useful starting point, sanctioned by both philosophical tradition and ordinary usage, is to think about happiness as an emotion. Happiness and cognate states such as joy and gladness have been reckoned among the "passions" by philosophers since Aquinas; today we would naturally call them "emotions" or "feelings". Some philosophers use the word "happiness" to translate *eudaimonia*, which is a condition of life rather than an emotion, but this semi-technical usage is presumably not what the designers and subjects of happiness surveys have in mind. Other philosophers define happiness as "satisfaction with one's life as a whole". This definition comes closer to ordinary usage and is embedded in the design of many happiness surveys. However, it is not inconsistent with thinking about happiness as an emotion. Emotions can, as I argue below, involve judgements and persist over many months or years. Besides, *mere* satisfaction with one's life, without any accompanying feelings – a "cold-blooded and dispassionate judicial sentence", as William James put it – is unrecognisable as happiness (James, 1884, p. 194). So in what follows, I shall respect ordinary usage and treat happiness as an emotion.

Emotions have three features that make them vulnerable to introspective error: they are dispositional, they are intentional, and they are publically manifest. Let me take these in turn.

Emotions are dispositional. To say that John is in love with Mary or jealous of his boss is not usually to say that he is currently feeling a certain way about Mary or his

boss but rather that he is generally disposed to feel that way, and to act accordingly. An emotion, writes Peter Goldie, "involves dispositions, including dispositions to experience further emotional episodes, to have further thoughts and feelings, and to behave in certain ways" (Goldie, 2000, pp. 12-13). This obvious point is sometimes overlooked in the psychological literature on the emotions, where (perhaps for reasons of experimental convenience) emotions are often identified with short episodes of intense feeling. But we have no reason to accept this restriction.

Happiness and unhappiness also involve dispositions. The sentence "John is happy" can, it is true, sometimes be used to make a statement about John's current state of mind. ("John is happy. He took Ecstasy an hour ago.") But unless it is qualified in some such way, it is more naturally understood in a dispositional sense, as a statement about John's standing tendency to feel and act in certain ways. It tells us that John is often in a good mood, that he smiles readily, that he is likely to confront misfortune without despair, and so forth. Participants in happiness surveys are usually asked how happy they are "in general" or "taking their lives as a whole", implying that what is at issue is dispositional, not occurrent happiness.

The dispositionality of happiness is one reason why it, like other emotions, is not reliably accessible to introspection. For it is a well-known fact that people are often very bad judges of their own dispositions. They tend to ascribe a false permanence to their current feelings, forgetting how often they have felt differently in the past and how readily they may feel differently in the future. This is particularly true of the young, who have yet to learn the fickleness of human passions. Pushkin's novel in verse, *Eugene Onegin*, provides us with a nice example. Olga feels genuinely and acutely unhappy over the death of her betrothed, Lensky, but we cannot call her deeply unhappy because we know – Pushkin tells us so – that her native cheerfulness will soon reassert itself. Presented with a happiness questionnaire, Olga might well respond in a way that a perceptive onlooker would regard as unduly pessimistic.

Some have denied that happiness is dispositional. As mentioned above, Daniel Kahneman holds that happiness in an interval is simply the sum of happiness at moments within that interval.[1] This seems to me implausible. Let us suppose that Olga has a sister who is also in love with Lensky. A week after Lensky's death, the two women are equally unhappy, yet whereas Olga will recover quickly her less flexible sister will remain grief-stricken for years. We would want to call Olga superficially, her sister deeply unhappy. Yet by hypothesis, the two women's lives contain, to date,

1. This view is also defended in (Feldman, 2010, p. 137)

an equal share of happy and unhappy feelings. Therefore happiness is not purely occurrant.[2]

And even if happiness is occurrent, it is still vulnerable to introspective error, though from a different direction. On an occurrant understanding of happiness, the question "how happy are you in general?" can only mean "how happy have you been, moment by moment, within some past interval?" And answers to this latter question are open to the doubts raised by Kahneman about the reliability of emotional memory. These doubts might be assuaged in the way that Kahneman suggests, by monitoring happiness at regular intervals and summing the results, but the difficulties of this procedure are such that it has never been implemented on any scale. Besides, the mere act of reporting regularly upon one's happiness might well be expected to have a depressing effect upon it. "Ask yourself whether you are happy, and you cease to be so" wrote John Stuart Mill (Mill, 1924, p. 120).[3]

Emotions are intentional. Most recent literature on the emotions has emphasised their intentionality – their directedness towards an object.[4] Generally speaking, we don't just feel frightened or angry; we feel frightened *of* x or angry *about* y. Where this object is a complex state of affairs, emotion furthermore implies belief – belief, at a minimum, that the state of affairs obtains. If I am angry that the local hospital is in a mess, I presumably believe that the local hospital is in a mess.

The conceptual connection between emotion and belief gives us yet another reason to doubt the authority of emotional self-reports. It is a familiar if puzzling fact that people can *believe* they believe things that they do not really believe: this is the phenomenon of insincerity or bad faith. And if beliefs can be insincere or in bad faith, emotions based on those beliefs can also be insincere or in bad faith. If we are sceptical of condemnations of private education from people who send their sons to Eton, we can also be sceptical when such people express *outrage* over private education. We needn't deny that the "outrage" feels real to those experiencing it, or that it is accompanied by its usual behavioural manifestations – shrill voices, knotted brows etc. What makes it insincere is the absence of the connections that should normally exist between it and the general course of life. The case is not unlike that of the skilled actor who works himself up into a frenzy of indignation over some purely fictional wrongdoing.

2. For a more detailed defence of the dispositionality of happiness, see (Haybron, 2008, p. 69)
3. (Feldman, 2010, pp. 98-104) also discusses the issue.
4. See for instance, among many others, (Solomon, 1984), (Goldie, 2000), (Nussbaum, 2001).

Happiness, too, has an essential connection to beliefs about the world. Generally speaking, one is not just happy, but happy that such-and-such is the case. This understanding of happiness as intentional faces two apparent counter-examples. First, one can be "in a happy mood" without being happy about anything in particular. However, as Peter Goldie has convincingly argued, "a mood involves feeling towards an object just as much as does an emotion, although ... what the feeling is directed towards will be less specific in the case of a mood." (Goldie, 2000, p. 143) When I'm in a happy mood I'm not happy about this or that but about many things or things in general. I warm to the dull old gentleman on the bus; I forgive the insult I received this morning; I may even, if I'm metaphysically inclined, start looking on the world as intrinsically just and beautiful. "The world of the happy is quite another than that of the unhappy" wrote Wittgenstein in the *Tractatus* (Wittgenstein, 1922, 6.43).

Second, one can be said to be simply happy, without further qualification. "How's Jane these days?" "She's very happy." Such "all-in-all" happiness looks, on the face of it, non-intentional. But it cannot really be so. If Jane is not happy about anything, unhappy about many things, and has no tendency to be in a happy mood, it makes no sense to call her happy. All-in-all happiness is logically tied to happiness about specific things or things in general, which is not to say that it can be derived from them by means of some algorithm.

If happiness is essentially grounded in beliefs about the world, it can share in the insincerity of those beliefs. This kind of insincerity is, I suspect, quite common. Think of the man who, as part of a positive thinking course, is required to repeat the mantra "every day in every way I'm getting a little bit better." After a while, he comes to affirm this thought quite spontaneously, though in sober moments he acknowledges that it is not really true. He is happy that his life is getting better. He scores himself 8 out of 10 on happiness questionnaires. But is he really happy, if his considered belief is that his life is not getting better at all? It is at least plausible to suggest that the answer is "no".

Emotions are publically manifest. If "fear" were the name of a purely private sensation, with only a contingent relation to public circumstances and behaviour, it would make good sense to ascribe it to a man who neither displays nor has cause to display fear. But such an ascription makes no sense. It merely suggests a misunderstanding of what the word "fear" means. The point is familiar from Wittgenstein. Words referring to inner states stand in need of outward criteria of application. We manifest understanding of the word "fear" by using it correctly, on the basis of our knowledge of

the situations, gestures and actions that typically accompany fear. These situations, gestures and actions constitute the frame of reference within which alone "fear" has meaning.

One implication of this is that it is possible for an individual to be mistaken as to whether or not he feels an emotion. John may think and say that he respects his colleague Sarah, but if his behaviour towards her is not of a respectful kind – if he continually ignores her and puts her down, say – we might reasonably be doubtful. Perhaps John has managed to conceal from himself his real feeling, which is one of contempt. Were emotions purely phenomenal states, such an ascription would be puzzling – like ascribing to John a pain that he does not feel. But if they are criterially connected to outward behaviour, it is perfectly intelligible. People are often blind to the emotional tenor of their actions. We might even agree with Proust that "it is only with the passions of others that we are ever really familiar, and what we come to discover about our own can only be learned from them." (Proust, 1992, p. 181)

Happiness, too, is criterially connected to public actions and circumstances, which gives us yet another reason to doubt the authority of first-person happiness reports. A woman who says that she is happy but whose actions and circumstances suggest otherwise is not self-evidently credible. Perhaps she is "in denial". Perhaps she is a Stoic philosopher with an unusual understanding of happiness. Of course, an individual's assertion that he is happy may be among the grounds for ascribing happiness to him: this is, after all, one common way in which happiness manifests itself. But it is not the only way. Where verbal and non-verbal manifestations of happiness conflict, only the particularities of the individual case can tell us which to trust.[5]

It is, then, simply not true that "if people say they are happy then they *are* happy". Individuals are not authoritative judges of their own happiness. They can be mistaken.

ARE PEOPLE IN GENERAL RELIABLE JUDGES OF THEIR OWN HAPPINESS?

But perhaps advocates of the survey method needn't insist on the infallibility of happiness self-reports. All they need say is that such reports are, on average, accurate.

5. For a defence of a view similar to this, see (Kenny, 2006, pp. 135-148).

Sure, there will be errors of optimism and pessimism here and there. But given a large enough sample, these errors will "wash out".[6]

Supporters of this hypothesis take comfort from studies showing correlations between self-reported happiness and other measures associated with happiness. These measures are of three broad kinds: physiological, behavioural and circumstantial. On the physiological side, it has been shown that people who declare themselves happy tend also to have good immune systems and high levels of electrical activity in the left forebrain (see Layard, 2005, pp. 17-20). These look like significant correlations, but they invite the question: how do we know that such physiological indices themselves track *happiness*? The answer clearly cannot be that they track self-reported happiness, since that is the very thing in question. Such studies may strengthen our conviction that happiness self-reports latch onto something real, but they do not establish what this something is. For all we know, it might simply be a propensity to answer happiness questionnaires optimistically.

Other studies show a correlation between self-reported happiness and the actions and circumstances associated with happiness. Andrew Oswald and Stephen Wu have established a correlation between quality of life in U.S. states, as measured by sunshine hours, commuting times, crime figures etc., and the self-reported happiness of their inhabitants. (Oswald and Wu, 2010. Adjusting for age and wealth, New York comes out bottom on both counts.) Other studies have shown that people who rate themselves happier also tend to smile more frequently (see Diener and Suh, 1999, p. 437).

If non-verbal actions and circumstances are, as I have claimed, among the criteria of happiness, then these studies must indeed increase our confidence in happiness self-reports. But they also raise an important epistemological puzzle. Happiness self-reports are not, I have said, self-evidently authoritative. They require external validation. But in seeking such validation, do we not presuppose a measure of happiness independent of self-reports? How then can happiness surveys tell us anything new? Either they tally with prior estimates of human happiness, in which case they seem to be redundant, or else they do not, in which case they seem to be flawed. Their function, it appears, is essentially ceremonial: it is to bestow the blessings of social science on the deliverances of common sense.

6. Daniel Haybron defends the validity of happiness surveys along these lines, despite his doubts concerning the reliability of happiness self-reports in the individual case. See (Haybron, 2007, p. 412)

This is too strong a conclusion.[7] In science, the fact that measure A must be validated against measure B does not mean that it cannot in turn refine, extend and, on occasion, correct B. In fact, this is a common occurrence. Hasok Chang has written a fascinating book on the history of modern thermometry, showing how each new measure of temperature has had to be validated with reference to earlier, less sophisticated measures. We confront, he writes, "the paradoxical situation in which the derivative standard corrects the prior standard in which it is grounded" (Chang, 2004, p. 44). How is this possible?

Chang identifies two principles governing the advance of temperature measurement. The first is an "imperative of progress" (Chang, 2004, p. 44). Progress here has a number of aspects. A new thermometer can be superior to existing instruments in accuracy (it registers smaller intervals), in range (it measures temperatures above and below what was previously possible) or in reliability (it is less prone to error). A mercury thermometer is superior to a human hand – the primordial thermometer – in all three respects. It can register finer gradations of temperature. It can measure temperatures above the point at which a hand burns and below the point at which it goes numb. And it is not prone to the illusion that the same object is cold (to a hot hand) and hot (to a cold hand).

But alongside the imperative of progress Chang also posits a "principle of respect" (Chang, 2004, p. 43). This states that a new measure must, by and large, agree with the measure it replaces. Otherwise, we have no grounds for saying that it is measuring the same thing, or indeed anything at all. The principle of respect limits the degree to which a new measure can correct an old one. A thermometer may on occasion override the evidence of our unaided senses, but if it does so too often, it is no longer clear that it is measuring *temperature*.

Chang's work provides a useful framework for thinking about the achievements and limits of happiness surveys. Like thermometers, happiness surveys lend mathematical precision to the rough and ready verdicts of intuition. Common sense tells us that health, love and honour all contribute to happiness, but it does not tell us how *much* they contribute. Surveys inform us that the effect of unemployment on happiness is 20 per cent greater than that of divorce (Layard, 2005, p. 64).[8] They tell us that men adapt better to divorce than do women (Layard, 2005, p. 66). And they

7. In a previous discussion of the subject, I myself advanced this strong conclusion. I now think the matter is more complex. See (Skidelsky, 2012, p. 112)

8. Divorce leads to a fall in happiness of 5 points on a 10-100 point scale. Unemployment leads to a fall in happiness of 6 points.

tell us that no one adapts well to the irritations of background noise (Frederick and Loewenstein, 1999, p. 311).

These are not banal findings. They tell us something new. Notice, though, that they do not *contradict* our prior understanding of what makes people happy. They merely refine it, in the same way that a thermometer with 0.1°C intervals refines one with 1°C intervals. Might happiness surveys do more than this? Might they not just refine but substantially revise our common-sense understanding of the conditions of happiness? I suspect not, because our only ground for trusting happiness surveys is that they by and large *agree* with common sense. Were they to disagree significantly, we would not trust them. We would suspect some computational error, or question the sincerity, self-knowledge or linguistic competence of the participants. Here Chang's "principle of respect" comes into play. A new measure can contradict its predecessor only on occasion, as in the case of the thermometer versus the hot and cold hand. If it contradicts it across the board, we have no reason to accept it.

The problem is not just hypothetical. Sometimes happiness surveys do yield strikingly counter-intuitive results. An example comes from the aforementioned paper by Oswald and Wu. "Although it is natural to be guided by formal survey data", they write (2010, p. 578),

> *it might be thought unusual that Louisiana – a state affected by Hurricane Katrina – comes so high in the state life-satisfaction league table. Various checks were done. It was found that Louisiana showed up strongly before Katrina and in a mental-health ranking done by Mental Health America and the Office Applied Studies of the U.S. Substance Abuse and Mental Health Services Administration .... Nonetheless, it is likely that Katrina altered the composition of this state – namely, those who were left were not a random sample of the population – so some caution in interpretation is called for about this state's ranked position, and that position may repay future statistical investigation.*

This is a revealing admission. While recognising that they should, in all consistency, "be guided by formal survey data", Oswald and Wu allow themselves to be swayed by what they intuitively know about the effects of hurricanes on happiness. When it comes to the crunch, they question the data; they do not revise their views on what makes people happy.

But what if "future statistical investigation" proved the Louisiana survey to be

representative after all? Should we conclude that natural disasters do not, contrary to widespread belief, detract from happiness? Not necessarily. We might first of all ask whether the Louisiana respondents were being honest, whether they had correctly understood the scale presented to them, and whether they were using the word "satisfied" with the same meaning as respondents elsewhere. We might also wonder – to return to the issue raised in the first half of this article – whether or not they knew their own minds. Ruling out these and other possibilities would require further, more detailed interviews, and perhaps also behavioural studies. We might in the end conclude that Katrina had indeed failed to dampen the good spirits of the citizens of Louisiana: this would be a case of a survey overturning prior expectations. But we would accept this conclusion only with some resistance, and only if accompanied by some intuitively plausible explanation of the psychological mechanism involved. (Perhaps disaster had awakened a spirit of community, as is said to have happened in London during the Blitz.)

Let me summarise. The criteria of happiness are, I have said, threefold: verbal, behavioural and circumstantial. All three categories are on a par, epistemologically speaking. There is no automatic presumption in favour of the first. Hence while happiness self-reports might occasionally be allowed to overrule the testimony of behaviour and circumstances, there is no necessity for them to do so. Indeed, Chang's principle of respect implies that it is *only* occasionally that self-reports can be allowed to overrule the testimony of behaviour and circumstances. Were they to do so too often, they would simply cease to be credible. In short, nothing that surveys might tell us can upset our common-sense conviction that health, love, freedom, security and respect all standardly contribute to happiness.

My conclusion, then, is a qualified endorsement of the survey method. Happiness surveys can add numerical precision to our prior understanding of the causes of human happiness. They can arrange those causes in order of importance, perhaps even assign cardinal values to them. These are significant achievements. But we cannot expect surveys to fundamentally revise our prior understanding of what makes people happy and unhappy. It is comforting to know that modern statistics confirm Solomon's dictum that a dinner of herbs with love is better than a stalled

ox with hatred.[9] But if they failed to confirm it, we would quite rightly side with Solomon, and not with the statistics.

Acknowledgements: I am grateful to the audience at the conference "Happiness and Well-Being" (Uehiro Centre, Oxford, 20-21 June, 2013), where I read an earlier version of this paper. I am particularly grateful to my colleague at Exeter University, Sabina Leonelli, who directed me to the work of Hasok Chang. I am also grateful to my two anonymous referees, whose useful criticisms of an earlier draft I have striven to address.

REFERENCES

Argyle, M., *The Psychology of Happiness* (London: Routledge, 1987).

Chang, H., *Inventing Temperature: Measurement and Scientific Progress* (Oxford: Oxford University Press, 2004).

Diener, E. and Eunkook M. S., "National Differences in Subjective Well-Being", in D. Kahneman, E. Diener and N. Schwarz, *Well-Being: The Foundations of Hedonistic Psychology* (New York: Russell Sage, 1999), pp. 434-450.

Feldman, F., *What is This Thing Called Happiness?* (Oxford: Oxford University Press, 2010).

Frederick, S. and George L., "Hedonic Adaptation", in D. Kahneman, E. Diener and N. Schwarz, *Well-Being: The Foundations of Hedonistic Psychology* (New York: Russell Sage, 1999).

Goldie, P., *The Emotions: A Philosophical Exploration* (Oxford: Clarendon Press, 2000).

Haybron, D. M., "Do We Know How Happy We Are? On Some Limits of Affective Introspection and Recall", *NOÜS* 41:3 (2007), pp. 394-428.

——— *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being* (Oxford: Oxford University Press, 2008).

Kahneman, D., "Objective Happiness", in *Well-Being: The Foundations of Hedonic Psychology,* eds. D. Kahneman, E. Diener and N. Schwarz (New York: Russell Sage Foundation, 1999), pp. 3-25.

Kenny, A. and C. Kenny, *Life, Liberty and the Pursuit of Utility* (St. Andrews Studies in Philosophy and Public Affairs, 2006).

Layard, R., *Happiness: Lessons from a New Science* (London: Penguin, 2005).

Mill, J. S., *Autobiography* (London: Oxford University Press, 1924).

Myers, D. G., "The Funds, Friends, and Faith of Happy People", *American Psychologist* 55/1 (2000), pp. 56-67.

9. Proverbs 15:17: "Better is a dinner of herbs where love is, than a stalled ox and hatred therewith."

Nussbaum, M., *Upheavals of Thought: The Intelligence of Emotions* (Cambridge: Cambridge University Press, 2001).

Oswald, A. J. and S. Wu, "Objective Confirmation of Subjective Measures of Human Well-Being: Evidence from the U.S.A.," *Science* 327 (29 Jan. 2010), pp. 576-79.

Proust, M., *Swan's Way,* trans. C. K. Scott Moncrieff and Terence Kilmartin (New York: Random House, 1992), p. 181.

Schwarz, N. and Fritz S., "Reports of Subjective Well-Being: Judgemental Processes and Their Methodological Implications", in *Well-Being: The Foundations of Hedonic Psychology,* eds. D. Kahneman, E. Diener and N. Schwarz (New York: Russell Sage Foundation, 1999), p. 61-84.

Skidelsky, R. and Skidelsky, E., *How Much is Enough: The Love of Money, and the Case for the Good Life* (London: Allen Lane, 2012).

Solomon, R., *The Passions: The Myth and Nature of Human Emotions* (New York: Doubleday, 1984).

Wierzbicka, A., "'Happiness' in Cross-Linguistic and Cross-Cultural Perspective", *Daedalus* 133/2 (2004), pp. 34-43.

James, W., "What is an Emotion", *Mind* 9 (1884), pp. 188-205.

Wittgenstein, L., *Tractatus Logico-Philosophicus* (London: Routledge and Kegan Paul, 1922)

Indirect Discrimination Is Not Necessarily Unjust

KASPER LIPPERT-RASMUSSEN

Aarhus University

ABSTRACT

This article argues that, as commonly understood, indirect discrimination is not necessarily unjust: 1) indirect discrimination involves the disadvantaging in relation to a particular benefit and such disadvantages are not unjust if the overall distribution of benefits and burdens is just; 2) indirect discrimination focuses on groups and group averages and ignores the distribution of harms and benefits within groups subjected to discrimination, but distributive justice is concerned with individuals; and 3) if indirect discrimination as such is unjust, strict egalitarianism has to be the correct account of distributive justice, but such egalitarianism appears vulnerable to the leveling down objection (whether decisively or not), and many theorists explicitly reject strict egalitarianism anyway. The last point threatens the position of liberals who oppose indirect discrimination but think significant inequalities can be just.

I. INTRODUCTION

In most Western countries many forms of direct discrimination are illegal. Employers can be fined and required to pay compensation if they reject applicants on grounds of race, gender, religion, or sexuality. Not only are such actions illegal. Most people consider direct discrimination on these grounds unjust across a wide range of contexts. I write "a wide range of context" and not "all contexts", because many do not believe it is unjust if, say, a film director "making a film about the lives of blacks

living in New York's Harlem" treats applicants differently on the basis of race (See Singer 1978, p.188).[1]

Initially, many hoped that once we got rid of direct discrimination, inequalities of race, and gender, and so on, would wither away, but clearly the legal prohibition of direct discrimination has not had this result. This is where indirect discrimination enters into the picture. The famous 1971 US Supreme Court ruling—*Griggs vs. Duke Power*—confirmed that a rule or practice can be illegal when it is "fair in form, but discriminatory in operation"—or, to put it differently, indirectly discriminatory (Fredman 2011, p. 178; Connolly 2011, p. 152; *Griggs v. Duke Power* 1971). In the case at hand an employer, Duke Power, "instituted requirements of high school education and satisfactory scores in an aptitude test as a condition of employment or transfer. The same test was applied to all candidates, but because black applicants had long received education in segregated schools, both requirements operated to disqualify black applicants at a substantially higher rate than whites" (Fredman 2011, p. 178). Since the relevant requirements were not needed to ensure satisfactory levels of performance, the company was ordered to abolish the requirement and to address the underrepresentation of Afro-Americans on its staff.

The 2009 Supreme Court ruling in *Ricci v. DeStefano* has to a large extent reversed the *Griggs vs. Duke Power* ruling. However, the idea that acts, practices and rules can be indirectly discriminatory, and therefore unjust, as a result of their differential effects, and in the absence of any intention to exclude members of any group, has had a huge impact; and many legal codes now recognize indirect discrimination as a prohibited category along with direct discrimination (*Ricci v. DeStefano* 2009; See also Selmi 2006). For instance, various European Court of Human Rights rulings have embraced the view that indirect discrimination falls under the *European Convention on Human Rights*. Also, EU Council directives mandate implementation of the principle of equal treatment irrespective of racial or ethnic origin in part through the prohibition of direct as well as indirect discrimination. (*DH v. Czech Republic* 2008; see also *Shanagan v. UK* 2014).[2]

1. For a similar claim in relation to so-called reaction qualifications in general, see (Wertheimer 1983, p.101; Alexander 1992, , pp. 173–176)

2. A similar legal stance is represented by Britain's *Equality Act 2010*, which prohibits direct as well as indirect discrimination in relation to certain "protected characteristics": "age; disability; gender reassignment; marriage and civil partnership; race; religion or belief; sex; sexual orientation." The Act states that "(1) A person (A) discriminates against another (B) if A applies to B a provision, criterion or practice which is discriminatory in relation to a relevant protected characteristic of B's. (2) For the purposes of subsection (1), a provision, criterion or practice is discriminatory in relation to a relevant protected characteristic of B's if—(a) A applies, or would apply, it to persons with whom B

While many liberals favour the legal prohibition of indirect discrimination, it raises a number of thorny issues. One is about the list of protected groups that most such prohibitions involve (see note 5). Why are the groups mentioned above on the list? Consider age. In some contexts rules that disadvantage certain age groups seem just. Rules of organ transplantation prioritizing the needs of young patients, who have enjoyed few worthwhile years of life, over those of older patients might be an example—e.g. rules of organ transplantation prioritizing the needs of young patients, who have enjoyed few worthwhile years of life, over those of older patients—seem just (Kappel and Sandøe 1992, pp. 297-316). Again, why are certain groups, such as the obese, or those on low-incomes, absent from the list?[3] Certainly, people with obesity or on a low-income are seriously disadvantaged by various rules and practices that seem—even are—fair in form.

These questions are hard to answer. However, they arise in connection with both direct and indirect discrimination, and my focus here is on questions specifically about the latter (Lippert-Rasmussen 2013, chapter 1). I begin, in Section II, by expounding an Altmanesque definition of indirect discrimination with the aim of presenting three core challenges to the view that indirect discrimination is unjust as such. Section III focuses on the distinction between local disadvantage, e.g. underrepresentation of certain groups among CEOs, and global disadvantage, e.g. disadvantage in terms of the overall of resources. This distinction gives rise to the local-global disadvantage dilemma: Either accounts of indirect discrimination concern the former, in which case indirect discrimination is not unjust as such, or they concern the latter, in which case they are radically revisionist. Section IV notes that mainstream theories of indirect discrimination determine disadvantage on the basis of group averages. The gives rise to the challenge from group averages: In the light of intragroup inequalities, indirect discrimination is not always preferable, justice-wise, to its absence. Section V shifts the focus from disadvantage to disproportionality—both essential components in indirect discrimination—and distinguishes between two interpretations thereof: One that compares inequalities between groups under situations with and without indirect discrimination—the relativized view—and one

does not share the characteristic, (b) it puts, or would put, persons with whom B shares the characteristic at a particular disadvantage when compared with persons with whom B does not share it, (c) it puts, or would put, B at that disadvantage, and (d) A cannot show it to be a proportionate means of achieving a legitimate aim." (*Equality Act 2010*), cf. (Connolly 2011, pp. *55–77*).

3. For discrimination against obese people, see (Harnett 1992-1993). For income discrimination, see (Lippert-Rasmussen 2013).

that compares how well off the group being subjected to indirect discrimination is under situations with and without indirect discrimination against it—the absolute view. Section VI shows that only the former view fits standard conceptions of indirect discrimination. However, this implies—and that is my third and final challenge—that the view that indirect discrimination as such is unjust is vulnerable to the so-called levelling down objection and, thus, that to endorse this view one has to reject this objection. I conjecture that many who find indirect discrimination unjust will find this an unwelcome implication of their view. After all, it is commonly assumed that one can consistently oppose indirect discrimination without subscribing to strict egalitarianism. Section VII responds to three objections to my levelling down challenge and makes some cautious remarks about its limitations. Section VIII concludes by exploring the practical implications of the views defended here, i.e., that because indirect discrimination is not unjust as such acts that indirectly generate group disadvantages need not be unjust and, thus, might be such that they should be legally permitted.

Political philosophers have paid surprisingly little attention to the question *why* discrimination is unjust compared to other political charged questions such as "What makes wars just?" and "Should abortion be legal?" In fact, I do not think that there is a reasonably well established, or well-expounded view of what makes discrimination unjust (when it is). Accordingly, this article should not be seen as a refutation of such a view, but more as an important new step into under-theorized territory in political philosophy. That being said, the view that discrimination as such, and by implication indirect discrimination which after all is a species of discrimination, is unjust is common. For instance, James W. Nickel writes: "Discrimination is morally wrong because its premise that one group is more worthy than another is insulting to its victims, because it harms its victims by reducing their self-esteem and opportunities, and because it is unfair" (Nickel 2000, p.214). Similarly, Lena Halldenius uses the term "discrimination" such that "[w]hen an action has been correctly described as an instance of discrimination, it has at the same time been correctly described as unfair" (Halldenius 2005, p. 456). [4] In my view, the assumption that discrimination as such is unjust deserves closer scrutiny. This is true of direct as well as indirect discrimination, but, as already noted, here I restrict my attention to indirect discrimination and it is more plausible to deny that indirect discrimination, as opposed to direct discrimination, is unjust as such, because the latter involves treatment that is unfair

4. I take it that if something is unfair it involves a violation of comparative justice.

or involves objectionable mental states irrespective of its consequences (See however Lippert-Rasmussen 2013, pp. 103-189).

While it is not really necessary to mount my three core challenges, throughout this article I shall assume that if a certain act is unjust that constitutes *a* reason for the moral desirability of the act being legally prohibited. This assumption is quite weak and is acceptable to a wide range of theorists. First, it does not rule out there being non-justice based, potentially overriding, reasons for or against legal prohibitions of acts, e.g. that they promote or reduce general welfare. Second, on many views injustice is cashed out in terms of violation of rights—in the case of discrimination: *human* rights—and it is commonly assumed that the law ought to prohibit (human) rights violations. Finally, legal moralists believe that the fact that an action is morally wrong is *a* reason to prohibit it. While some endorse legal moralism, many reject it, but even most of those, who do, accept that the subset of morally wrongful acts that involve injustice ought, morally speaking, to be legally prohibited.

II. INDIRECT DISCRIMINATION DEFINED

To determine whether indirect discrimination as such (henceforth I take this qualification for granted) is unjust we need to know what it amounts to, since, presumably, if it is unjust *as such* (henceforth I take this qualification for granted), it is unjust in virtue of features that *necessarily* belong to it.[5] In particular, we need to have a clear view of the respects in which indirect discrimination differs from direct discrimination. In an encyclopedia entry on discrimination, Andrew Altman rightly notes that there is no agreed test, or criterion, of indirect discrimination (Altman 2011). Still, drawing on Altman's work I propose the following definition, one that fits a number of existing characterizations of indirect discrimination quite well (Cf. (Halldenius 2005, p. 459):

> *A policy, practice or act is indirectly discriminatory against a certain group if, and only if: 1) it neither explicitly targets nor is intended to disadvantage members of the group (the no-intention condition); 2) it disadvantages members of the group (the dis-*

5. For some readers it may be helpful to note that I am exploring whether indirect discrimination is pro tanto unjust.

advantage condition); and 3) the relevant disadvantages are disproportionate (the disproportionality condition).[6]

All three conditions point to differences between direct and indirect discrimination. The no-intention condition captures the core difference—the idea being that a company, say, could indirectly discriminate against women even if it is neither explicitly targeting them (e.g. in job advertisements that invite applications from men only) nor intending to disadvantage them. (There is a different and non-intention related sense of "indirect", which should be distinguished from the sense of "indirect" I expound here: this is the sense in which using a certain proxy (e.g. being taller than 1.85 meters) for pursuing one's aim (excluding women) is indirect. Discrimination that is indirect in this sense is direct discrimination in my sense.)

The disadvantage condition also captures a difference between direct and indirect discrimination. To directly discriminate one has to treat the *discriminatee* of one's actions disadvantageously in some way. However, in some circumstances one can do this without the outcome of one's actions actually being disadvantageous to the discriminatee. Suppose a homophobic employer initially decides to hire a straight applicant rather than a better qualified gay applicant, but is then forced to offer the job to the latter because the former withdraws his application. The gay applicant was subjected to direct discrimination—the employer initially decided not to hire him on account of his sexuality—even if, as it so happened, the relevant outcome was not harmful for him. More generally, while indirect discrimination is tied to the outcome of the allegedly discriminatory process, direct discrimination requires only that a person be subjected to disadvantageous treatment. (Here I set aside here outcome-focused conceptions of direct discrimination according to which cases such as the one I described above involve attempted, but unsuccessful, direct discrimination.) (Lippert- Rasmussen 2013, p.18; Gardner 1996; Connolly 2011, p. 155)

The disproportionality condition reveals a third difference between direct and indirect discrimination, for neither it nor any similar condition must be satisfied in cases of direct discrimination. Suppose there is some morally good reason to engage

6. Altman's definition implies that it is only socially salient groups that can be subjected to indirect discrimination. I omit this part of his definition, because, as noted in Section I in relation to the issue of the nature of protected groups, my focus is on issues that pertain specifically to indirect discrimination, as opposed to discrimination in general; but see (Lippert-Rasmussen, 2013, chapter 1). Sometimes people use a moralized concept of indirect discrimination such that if something is indirect discrimination, it is by definition unjust (or morally unjustified). I set aside this concept here. The discussion I present can be read as showing that much of what people who employ the moralized concept identify as indirect discrimination does not fall under their concept.

in direct discrimination. For example, we are in a country with a conservative, sexist majority that will predictably descend into civil war unless the established church directly discriminates against women when appointing people to religious offices. The interest in avoiding civil war morally outweighs the interest in sexual equality in the process of making church appointments. Here women are directly discriminated against when they are not hired for the relevant positions. Yet, because the case does not satisfy the disproportionality condition, a similar, but indirect case would not involve discrimination.

While this account of indirect discrimination can be improved upon in various ways, it suffices for our purposes, (Lippert-Rasmussen 2013, chapter 2) and I want now to tackle some issues raised by the question whether indirect discrimination is unjust. In doing so, I shall disregard the no-intention condition and focus on conditions 2) and 3). It is possible that indirect discrimination is unjust because it satisfies the disadvantage or the disproportionality condition. However, it cannot be unjust, because it neither explicitly targets, nor is intended to disadvantage, members of a certain group. After all, if targeting or intending to disadvantage makes a moral difference, justice-wise, it makes a difference to the worse, not the better.

III. LOCAL V. GLOBAL DISADVANTAGE

I begin with the disadvantage condition—the notion that indirectly discriminatory practices always disadvantage the group discriminated against. This condition is in need of clarification in two dimensions at least, and in ways that challenge the view that indirect discrimination is unjust. First (I will come to the second clarification in Section IV), a practice may disadvantage members of a certain group locally, or globally, as it were. If, on the one hand, disadvantage is understood locally, our concept of indirect discrimination is non-revisionist, but indirect discrimination is not unjust. If, on the other hand, disadvantage is construed globally, indirect discrimination is possibly unjust, but the emerging notion of indirect discrimination is also highly revisionist. This is the local-global disadvantage dilemma.

To see what the distinction between local and global disadvantage amounts to, imagine that language tests used by humanities faculties to select students tend to result in the admission of fewer immigrants. However, instead of being admitted to the humanities they seek admission at law schools, medical schools, engineering schools, and the like, where, as a result they are overrepresented. Suppose also that

as a result they end up living lives which are better than the lives of non-immigrants. Here, a formally neutral rule disadvantages immigrants locally: they find it harder to meet the language test and struggle to gain admission to the humanities faculty. But the same rule advantages the immigrants globally: they end up being better off overall than other members of society. Whatever objectionable features the relevant admission rule has in virtue of its impact on global distribution, injustice to immigrants cannot be counted among them.

When a rule or practice is criticized as indirectly discriminatory, the focus is on local, not global, disadvantage—e.g. the disadvantage reflected in the fact that women are underrepresented among professors or CEOs. This may reflect our tendency, when raising complaints about indirect discrimination, to become exercised by local disadvantages that we take to contribute to a connected global disadvantage. This is why, presumably, although there are some rules and practices that place men at a local disadvantage (think of parental access to children following divorce), it is rare to hear of indirect discrimination against men (See, however, Sullivan 2004).

In the moral assessment of indirect discrimination the distinction between local and global disadvantage becomes important. Many would say that justice is concerned with the distribution of global benefits and burdens. On this view, the fact that some people are better off than others in some particular dimension—say, they have a higher income—can be counterbalanced by the fact that they are worse off than others in another dimension—they have longer working hours and less autonomy in their jobs. Undoubtedly, there is something right in the view that justice is concerned with the distribution of global benefits and burdens; it would be odd to hold that it makes no difference, from the point of view of justice, whether local disadvantages counterbalance or accentuate one another. Against this view, it might be argued that it would be odd for an indirectly discriminating employer to get off the law's hook simply because members of the group which she disadvantages, say, in terms of employment are advantaged in terms of other local goods, but in ways that are beyond this employer's control. However, insofar as disadvantaged groups are identified not relative to each individual employer but, say, relative to the job market as such, it is any case true that individual employers are held responsible in part on the basis of facts that they do not control.

Some people, notably Michael Walzer, have defended the view that there are different spheres of justice, and that justice requires the goods within each sphere to be distributed according to criteria reflecting the nature of the relevant goods. For

example, medical services should be distributed according to need, and places at universities according to merit (Walzer 1983).[7] On a hybrid, Walzerian view where justice requires each good to be distributed according to its cultural meaning and equality of global advantage, indirect discrimination could be unjust because it results in local disadvantage. Obviously, alternative hybrid views of the way local disadvantage matters can be envisaged, but Walzer's view is certainly the best known.

In response to the Walzerian position here, I note, first, that complaints about indirect discrimination often relate to disadvantages which, even on Walzer's view, involve local disadvantage within a certain sphere. If, for instance, certain rules and practices lead to worse health outcomes vis-à-vis a particular disease for women from a Walzerian perspective, this would qualify as a local inequality in the treatment of a particular medical need, and yet it is compatible with the sphere of health as a whole being just in the sense that, globally speaking, health care is distributed according to need overall. Second, on Walzer's view the social meaning of many goods implies they should not be distributed equally—e.g. admission to university should be based on merit. Accordingly, on Walzer's view one group might be worse off than others in terms of the distribution of a particular good without this distribution violating the social meaning of the good, in which case it could not involve injustice, let alone unjust, indirect discrimination. Hence, one cannot build an account of the injustice of indirect discrimination on Walzer's theory of justice. This completes my presentation of the local-global disadvantage dilemma.

IV. GROUP AVERAGES AND INTRAGROUP INEQUALITY

Let us now turn to the second dimension in which the notion of group disadvantage needs to be clarified. The basic issue here is that members of a group may be affected differentially by rules that, on average, (dis)advantage members of the group. Consider a test used to appoint senior managers which places a premium on being assertive, and assume it has following features. On average, women tend to score less well than men on it. Accordingly, despite equal numbers of men and women applying, more men than women are hired. Women and men vary in terms of how assertive they are. Some women are more assertive than most men, and some men are less assertive than most women. So, while it might be true that the test in question

7. For a reply defending the view that it is the distribution of global benefits and harms that matters, see (Arneson 1995)

disadvantages the women, the sub-group of especially assertive women may actually benefit from the rule (some would not have been hired had the test not been used) and the sub-group of especially unassertive men are harmed by it (some would have been hired had the test not been used). Given these features, the charge that the test indirectly discriminates against women and in favour of men seems insufficiently specific. Why not say that it indirectly discriminates against the sub-groups of unassertive people, men and women?[8] In itself this is an interesting question, but even if it can be answered in a principled way, there is another more worrying problem.

A rule, which, on average, disadvantages members of a certain group relative to another group, may in fact benefit most members of the group modestly provided that a few members are harmed a great deal (Cf. Doyle 2007). It may also be true that the few members who are seriously harmed by the rule are much better off than the rest in the absence of the rule. By way of illustration:

	5% best off men	All other men	Men average	5% best off women	All other women	Women average
Benefits under Rule I	490	90	110	130	110	111
Benefits under Rule II	100	100	100	120	120	120

On average, Rule II makes men worse off, but it also reduces the inequality between most men and most women, and it reduces male intragroup inequality since the harm it causes relative to Rule I falls on the 5% best off men.

Again, in response to these facts it is seriously inadequate simply to say that Rule II indirectly discriminates against men—for two reasons. First, in the absence of Rule II most men would be even worse off relative to members of other groups, so, given a plausible measure of the injustice of overall inequality, Rule II may in fact reduce unjust intergroup inequality (Temkin 1993, pp. 19 - 52). Hence, if we feel indirect discrimination is unacceptable because we find group inequality objectionable, this is a case of indirect discrimination we should not object to. Second, Rule II reduces

8. This example brings out the core issue of intersectionality and discrimination: that, at one and the same time, individuals might be discriminated against and in favor of in many different capacities; see (Crenshaw 1998)

intragroup inequality between men, and if we think that justice is more concerned with the plight of badly off men than with that of privileged men, it is not clear that we should prefer Rule I over Rule II from the point of view of justice, small benefits to many worse off people may outweigh substantial harms to a few better off people even if the total sum of benefits is greater in the outcome that favours the better off. Hence, if "indirect discrimination" picks out an injustice (or, at any rate, a prima facie injustice), then, despite the fact that Rule II makes men worse off on average it should not qualify as a case of indirect discrimination. This is the challenge from group averages.

Admittedly, this challenge assumes, first, that views of justice that focus on inequalities between groups ignore intragroup inequalities between individuals by favouring some trade-offs of greater inequality between individuals for less inequality between groups and, second, that this renders such views implausible (Holtug and Lippert- Rasmussen 2007, pp. 6 – 7). I find both claims plausible. Indeed, in my example Rule II seems to involve less objectionable inequality than Rule I despite that, on a view that focuses on group averages, it is the former which involves more indirect discrimination.

Admittedly, if disadvantages tend to cluster, the gap between local and global disadvantage explored in Section III will rarely arise (Wolff and De-Shalit 2007). Similarly, if it almost never turns out that on average a rule disadvantages, say, an oppressed minority even though most of its members actually are better off living under the rule than they would be in its absence, the challenge from group averages will almost never be a practical problem. (Of course, in the *Griggs v. Duke Power* case African-Americans with a high school degree were in one respect better off with Duke Power's rules of promotion than they were without it, since they faced no competition from fellow African-Americans without a high school degree.) On these assumptions, it is often best from the perspective of a political reformer to disregard such cases.[9] However, if we look at indirect discrimination from the perspective of the fundamental principles of justice—principles which are required to apply to all scenarios, and not merely to those that are actual or likely—we cannot ignore the local-global advantage dilemma and the challenge from group averages.

9. For an account of the difference between political advocacy and political philosophy, see (Cohen 2011, pp. 225–235.)

V. DISPROPORTIONATE MEANS: THE RELATIVIZED AND THE ABSOLUTE VIEW

To expound my third and final challenge, I first need to take a closer look at the disproportionality condition. Characterizations of indirect discrimination contain some such condition. For instance, the *Equality Act 2010* definition (see note 2) includes a disproportionality requirement, and in *Griggs v. Duke Power* the Supreme Court drew upon a proportionality clause to the effect that the exclusion of African-Americans had to be disproportionate in relation to job performance or business necessity.

Disproportionality is a relation between two items. One item—call it the bad item because it is this feature which invites the accusation that the rule or practice is unjustified—is disproportionate relative to another, which we might call the good item, because it can be called on in an effort to show that the rule or practice is justified, e.g., as in "The large amount of force used—a (very) bad item—was disproportionate to the relatively harmless threat thereby averted—a (minor) good item". To clarify the disproportionality condition we need to say a little more about these good and bad items.

Starting with the former, the first thing to note is that the use of the phrase "legitimate aim" in the *Equality Act 2010* can be misleading, in that it suggests that the good item is a certain sought for outcome, not the outcome itself. To see the difference, imagine the Supreme Court had instead found that Duke Power's high school requirement did indeed represent a business necessity, but also that the company operated this requirement neither with the aim of excluding African-Americans, nor in an effort to maximize business, but for some other reason that was legitimate. For example, the aim was to promote workplace harmony (which was not a business necessity) and the company believed, falsely, that a recruitment process ensuring that all members of senior staff had a high school degree would be one way of achieving this. Here there is no disproportionality, even though the company does not impose the high school requirement out of a concern for business necessity. What matters is that the requirement constitutes a business necessity. More generally, what matters is that there is some consequence (bankruptcy) of not applying the rule (or practice or policy) that justifies it, not whether the avoidance of this consequence is what motivates the agent whose decisions are being assessed for indirect discrimination.

The next question that arises in relation to the good item concerns the nature

of the relevant consequences—i.e. the currency of disadvantage. In *Griggs v. Duke Power* the consequences were couched in economic terms for the company in question. Impact on business was the key consideration. From a legal point of view, this narrow focus might make good sense. The assessment of the broader societal effects of a particular rule is difficult and, hence, to make law sensitive to such effects will make it hard for companies to know if they have infringed indirect discrimination laws. However, from a moral point of view it makes little sense to disregard these less easily quantified effects. For instance, a given admissions test may result in universities doing less well on narrow, university-related parameters (e.g. research output, donations, proportion of students graduating). At the same time, the use of this test rather than an alternative might generate much greater benefits for society generally (e.g. in terms of society being more tolerant and harmonious, and culturally and economically vibrant). In these circumstances, it would seem that, if we want our definition of indirect discrimination to include a disproportionality condition, these broader and beneficial consequences really ought to figure in the disproportionality at issue. Certainly, if the fact that a rule is indirectly discriminatory is a prima facie reason for thinking it is unjust, we should be willing to examine the proportionality of societal effects. Admittedly, doing so may raise more questions than it answers, because now we will now face tricky questions about how to assess a much broader range of consequences; there are many different suggestions as to what makes such consequences good, and as to how they should be weighed against one another. But these questions are not tied specifically to indirect discrimination. They are tied up with much more general issues in moral philosophy. Having flagged them, I will move on.

Let us now turn to the other of the two items in the disproportionality condition: the bad item. Two views here merit examination. The first is that a group is disadvantaged by a rule if, and only if, the inequality between this group and groups with which it is to be compared is greater with the rule than it would be in some relevant alternative situation without it. The second is that a group is disadvantaged by a rule if, and only if, this group would have been better off in some relevant alternative situation without it. Let us call the first view the group-relative (or simply relativized) view, and the second the absolute view.

To see the difference, consider a company that has a choice between two hiring policies. One involves hiring on the basis of qualifications only. The other involves hiring on the basis of qualifications on condition that the group of appointees faith-

fully reflects the make-up of society as a whole in respect of the protected groups (Spanish-speaking as opposed to English-speaking people, let us say, and suppose that these two groups have the same number of members). It turns out that the second policy results in the company not always hiring best-qualified applicants. This means the company will do less well commercially and end up hiring fewer people. (It is often argued that representational aims improve the competitiveness of a company. I want to steer clear of this empirical issue to address the normative issue of whether indirect discrimination is unjust if it reduces competitiveness in a certain way.) Moreover, in fact, the company will hire more people from any of the protected groups if it always hires the best qualified people than it would if it were to apply the second hiring policy. The situation is as follows:

	Number of English-speaking people hired	Number of Spanish-speaking people hired	Percentage of those hired who are women
Hiring policy 1	400	200	Approx. 33%
Hiring policy 2	180	180	50%

Suppose, finally, that we do not have to worry about consequences like objectively demeaning messages, e.g., it is not the case that severe underrepresentation of one group will objectively signify that members of the underrepresented group are inferior and deserve less concern and respect than others (Hellman 2008). On the relative view of the disproportionality condition, the first hiring policy may well be indirectly discriminatory, but the second policy is not so. On the absolute view, the first hiring policy is not indirectly discriminatory, while the second policy is. Indeed it might qualify as a policy that indirectly discriminates against English- and Spanish-speaking people. This implication is strikingly revisionist. It illustrates the general idea that, in principle, inequality is capable of making members of the worse off group better off than they would be under equality. John Rawls appealed to this general idea in defending his renowned "difference principle" of justice. The principle says that, subject to certain constraints, a just society is one in which the worst off in society are as well off as possible. From this it follows that inequalities are tolerable when, and to the extent that, they are required to make the worst off better off (Rawls 1971, pp. 302-303).

Interestingly, the general idea has gone largely unnoticed in discussions of indirect discrimination. Most who work in this area simply assumes a relative view of disadvantage. Thus it is common to find writers inferring, from the underrepresentation of a group, that this group is (probably) being subjected to indirect, if not direct, discrimination (Craig 2007, p.122). Because this inference is clearly invalid on the absolute view, one charitable interpretation of the views of those who make this inference is that they are wedded to the relativized view of disadvantage.

VI. THE LEVELLING DOWN OBJECTION AND INDIRECT DISCRIMINATION

How does the difference between the relativized and absolute view of disadvantage bear on the claim that indirect discrimination is unjust? In answering this question, I want to bring in what is usually referred to as the "levelling down" objection to egalitarianism—an objection occupying a prominent place in recent discussions of distributive justice, but which has so far not drawn attention in discussions of discrimination. Suppose we subscribe to the following strict egalitarian view: it is "bad—unjust and unfair—" if some people are worse off than others (Temkin 1993, p.13; Parfit 1998, p.3). Apparently, this view implies that a situation in which half the population is at 150 units of whatever is the currency of justice (welfare, resources etc.) and the other half is at 120 is unjust compared to one in which everyone is at 100. On the strict egalitarian view, the second situation, in which everyone is worse off, seems to be in one way better, because less unjust, than the first, in which everyone is better off. It is in one way better because it is better in terms of justice. Yet, as Derek Parfit has argued, this looks implausible. How can one situation be in any way better, e.g. in terms of the justice of distribution, than another in any respect if it is in no respect better for anyone, Parfit asks? (Parfit 1998, p.3). Many have taken this question to lay down a powerful challenge to egalitarianism.[10] Moreover, it is even

10. Admittedly, Parfit (1998, pp. 6-7) seems to suggest that a certain form of egalitarianism—deontic egalitarianism according to which it is the way in which inequality is produced and not the unequal outcome in itself that is unjust—is not vulnerable to the levelling down objection. Elsewhere I have argued that deontic egalitarianism is so vulnerable if telic egalitarianism is (Lippert- Rasmussen 2007)). In any case, the very idea behind indirect discrimination is that its injustice lies in the unequal outcome it generates, not in the indirectly discriminatory acts themselves, which after all are "fair in form". Accordingly, I do not see how an objection to indirect discrimination could derive from deontic egalitarianism. More generally, I do not see which agent-relative restriction pertaining the "act itself", so to speak, that someone who indirectly discriminates can plausibly be said to violate. For instance, I do not think it is plausible that there is a deontological restriction against indirect discrimination where indirect discrimination makes people better off in the way explored in Section

more powerful because there is an alternative to what we have called strict egalitarianism which arguably possesses many of the attractions of that view yet appears to provide an answer to the levelling down objection: "prioritarianism" (See, however, Voorhoeve and Otsuka 2009). The defining idea of prioritarianism is that an equal sized benefit which accrues to a person who is better off on some absolute scale of well-being has less moral value than a benefit that accrues to a person who is worse off on such a scale of well-being. If benefits can be redistributed and the redistribution will not affect the overall sum of benefits, an equal distribution is best, according to prioritarianism. But the levelling down objection has no purchase. Benefits to people, however well off they are, have positive moral value, but that value decreases the better off these people are. The upshot is that a situation where some are worse off and none is better off can never be better in any respect than one in which some are better off and none is worse off.

Let us return now to the conception of indirect discrimination on which we take disproportionality to involve the imposition of relativized disadvantages on the discriminatee. This conception is vulnerable to a challenge similar to the levelling down objection. To see this, suppose that where indirect discrimination occurs the members of one ethnic group will end up with 150 and members of another group 120, and that where it is eliminated everyone ends up with 100. We can now see that if strict egalitarianism is vulnerable to the levelling down objection, the view that indirect discrimination is bad because it is unjust is vulnerable to something very similar. How can indirect discrimination be bad in any respect, e.g. in terms of justice, one might ask, when it is bad in no respect for anyone? This is the levelling down objection to the view that indirect discrimination as such in unjust.

VII. CHALLENGES

I now want to rebut three critical responses to the levelling down objection to the injustice of indirect discrimination presented in the previous section, although ultimately I will concede that the levelling-down challenge is not decisive. First, then, it might be suggested that if the present challenge is sound, a similar one can be mounted to direct discrimination. However, direct discrimination is indisputably unjust. Hence, the present challenge must contain an error. This response is problematic. Direct and indirect discrimination differ. Assuming that a purely outcome-

III, see (Kamm 2007, pp. 24, 170-173)

focused account of fairness is false, the latter is fair in form (recall the formulation in *Griggs v. Duke Power*), the former is not. Indirect discrimination, if it is unjust, is unjust in virtue of the disadvantages it involves for certain groups (Cf. (Cavanagh 2003, p.199; Alexander Forthcoming). If a company rejects African-Americans on grounds of race, it treats them unfairly and arguably this violates an agent-relative, deontological constraint.[11] However, when a company applies a certain test in a way that is indirectly discriminatory the application of the test in itself is not unfair—if it were, the case would probably involve direct discrimination instead. It is only where, in the circumstances, the application of a test disadvantages members of a certain group that injustice is perpetrated.

The second challenge says that the levelling down objection to indirect discrimination is irrelevant, because, as a matter of fact, it never happens that no one is better off in the absence of indirect discrimination and some are even better off. My response to this challenge has three parts. (a) Even if the empirical basis of the challenge is true, this does not render the levelling down objection irrelevant to my question about indirect discrimination. My question is whether indirect discrimination as such is unjust, and to explore this question we need to consider hypothetical cases as well as actual and likely ones. (b) If we ask a different question—namely, one about what we, as political agents trying to bring into being a world that is more just, should be focusing on—the factual assumption is relevant. If, as a matter of fact, the discriminatees in cases of indirect discrimination would be better off if we eliminated that discrimination, we have some reason to do the latter, and this remains so even if there are counterfactual circumstances where doing so would not benefit and perhaps even harm indirect discriminatees. (c) So far I have granted the objector the factual assumption that in all cases of indirect discrimination members of groups suffering it would be better off in its absence. I do not want to claim that this is false (but recall my remark about African-Americans, and Duke Power employees with high school degrees). However, I would point out that it is a very strong claim, and that backing it up with evidence is a daunting task. Moreover, as the debate about affirmative action shows, it is far from uncontroversial that eliminating indirect discrimination always benefits the discriminatee. Thus it has been claimed that a demeaning message is sent when the criteria of assessment are adjusted to favour otherwise underrepresented groups, and that in some cases the resulting message-related costs to such groups of

11. For some doubts about the view that the levelling down objection, mutatis mutandis, does not challenge deontological views of justice too, see (Lippert-Rasmussen 2007)

not simply hiring or admitting applicants on a straightforwardly meritocratic basis are too great (Strauss 1995; Adarand v. Pena 1995, p.241; See, however, Bowen and Bok 1998).

Third, in setting out the levelling down objection to indirect discrimination I imagined that the elimination of indirect discrimination would not be better in any way for the discriminatee. However, the disadvantaging of groups involved in indirect discrimination does symbolic harm that ceases to be done when the discrimination is prevented. Hence, it might be suggested that eliminating indirect discrimination is always better in one respect: it eradicates symbolic harm. It is true that some forms of indirect discrimination are symbolically loaded and clearly do affront, or are seen as an affront to, the affected groups. However, this is not true of all forms. The tests used to recruit Navy Seals indirectly discriminate against elderly people, yet they are not generally thought to harm them symbolically. And we can certainly imagine other cases where indirect discrimination would involve no (effective) symbolic harming—e.g. because members of disadvantaged groups remain unaware that they are being disadvantaged by the relevant rules or practices. This shows that indirect discrimination as such does not cause symbolic damage. Finally, even in cases where symbolic harm is involved, the harming might be outweighed, morally speaking, by other kinds of harm that would be done if the indirect discrimination were eliminated. It may so happen, for example, that lowering meritocratic standards would harm all of us, and visit harm on the discriminatees that outweighs the benefit they would enjoy when shielded from symbolic harm.

Since none of the three challenges above is convincing, I tentatively suggest that if the levelling down objection defeats strict egalitarianism, it defeats the view that indirect discrimination is unjust. Like egalitarianism, concern about indirect discrimination arises from uneasiness at the relative positions of different groups. This opens the door to the levelling down objection, because one can always imagine the relative positions being adjusted in a way that leaves everyone worse off in absolute terms than they were before the adjustment. As this formulation indicates, the feature of a view of distributive justice that makes it vulnerable to the levelling down objection is not that it claims that justice is equality, but that it claims that justice consists in a certain relation between people's distributive positions. A view according to which justice requires that no one is (or indeed one that requires that some are) significantly worse off than others is also vulnerable to the levelling down objection. For simplic-

ity I disregard this broader scope of the levelling down objection and simply focus on strict egalitarianism (See Lippert- Rasmussen Forthcoming).

Does the levelling down objection amount to a knockdown argument against the view that indirect discrimination is unjust? I am not sure. Strict egalitarians have developed responses to the levelling down objection which, suitably revised, can be deployed in a rearguard action here. Some have pointed out that values other than equality imply that one outcome can better than another, even if it is better for no one in any respect. In a retributivist perspective on criminal justice, for instance, a world in which criminals are justly punished might be assessed as better than one in which they are not, even if this is better for no one because punishment has no deterrent effect. Hence, if the Slogan that "One situation cannot be worse (or better) than another in any respect if there is no one for whom it is worse (or better) in any respect" (Temkin 1993, p.248) obliges us to reject a wide range of values other than equality, perhaps the intuitive cost of rejecting it is lower than the intuitive cost of rejecting equality, desert and all the other values that offend against the Slogan (Temkin 1993, p.261).

In a separate move, it has been argued that some of those who reject egalitarianism in response to the levelling down objection are not really in a position to do so (Persson 2008). Consider prioritarianism. On this view, if we transfer one unit of well-being from a well off person to a badly off person this will result in an increase in moral value. But where does this increase come from, one might ask? Ex hypothesi, the decrease in well-being experienced by the source is exactly as great as the increase in well-being experienced by the recipient of the well-being. Accordingly, the value the transfer brings into existence seems to be unconnected to the sums of well-being. This suggests that prioritarianism, like egalitarianism, is committed to the idea that values are not tied to well-being for individuals. Since prioritarianism is commonly adopted by those who press the levelling down objection, this reversal of the attack has considerable bite.

Finally, some egalitarians take a bullish stance: they insist that, because it follows straightforwardly from strict equality that a state in which everyone is worse, but equally well off, is in one respect—though not all things considered—better than one in which everyone is better off, though unequally so, this implication is something they were aware they were committed to all along. Accordingly, the levelling down objection cannot play the dialectical role of an objection—it does not point to an implausible implication to which egalitarians are committed and of which (until the

alleged objection was presented to them) they were unaware. Ironically, in view of Parfit's formulation of the levelling down objection, it is probable that more egalitarians now will take this attitude than would have done so 30 years ago.[12]

There is a huge literature on the levelling down objection, and my aim here is not to argue that it refutes the claim that indirect discrimination can be unjust (see Holtug 2011, pp. 181- 201). My arguments in Section III and Section IV suffice to support this conclusion. My main aim here is to argue that indirect discrimination is unjust only if a strict egalitarian view of justice is correct, and thus that the levelling down objection fails. Even if the result of the discussion is limited in this way, it is very significant. Many people who are not committed to strict egalitarianism think that discrimination, including indirect discrimination, is unjust. Indeed, one hallmark of contemporary liberal opposition to discrimination is the assumption that one can be opposed to discrimination without committing oneself to any form of strict egalitarianism. If the argument of this section is sound, this option is unavailable, at least in the case of indirect discrimination. Strict egalitarianism of a certain sort—i.e. one that focuses on socially salient groups—is tied to the view that indirect discrimination is unjust! Since many would not want to tie them together in this way, my claim that they stand and fall together forms my third challenge to the view that indirect discrimination as such is unjust.

VIII. CONCLUSION

If the reasoning behind the local-global disadvantage dilemma, the challenge from group averages, and the levelling down objection applied to indirect discrimination is sound, indirect discrimination is not necessarily unjust. Because I am not certain that the levelling down objection is successful, my own basis for asserting the main claim of this article derives from the first two reasons only. I put forward the last objection, in its non-conditional form, in an ad hominem way.

Some might find the claim that indirect discrimination is not necessarily unjust discomforting. For one thing, they might worry that anyone who is persuaded by them will have to approve the legalization of indirect discrimination and (more

12. Another response to the levelling down objection is to hold that equality is non-instrumentally valuable, but that it is so only on condition that it benefits someone: see (Mason 2001) Yet another response is that, necessarily, unjust inequality is bad for worse off people: see (Broome 1991, p. 165)

generally) stop worrying about it. Let us briefly consider whether these worries are warranted.

First, it does not follow from the fact that something is not unjust as such that it is not often unjust. Contracts between employers and employees are not unjust as such, but many are unjust all the same—e.g. because they involve exploitation of the vulnerability of the employee by the employer. This means that the arguments in this article are entirely compatible with the view (which I am neither affirming, nor denying, here) that many forms of indirect discrimination should be made unlawful, because they are unjust. Moreover, to the extent that one allows that something might be unlawful, not because it is unjust, but because its presence often indicates injustice elsewhere, one could also, consistently with what I have argued, hold that indirect discrimination should be unlawful.[13]

Second, even if indirect discrimination is neither unjust, nor even often unjust (or even sometimes unjust), we have to remember that justice is not the only moral value, and that other values might speak against indirect discrimination. For instance, the French Revolution famously acclaimed fraternity as well as liberty and equality. Arguably, fraternity is hard to realize in a society where some groups are seriously underrepresented in the most prestigious and well paid job categories (Anderson 2010, pp. 89–111; Cohen 2009, pp. 27-34). So even if such underrepresentation is not unjust, it might still be morally indefensible, all things considered, not to eliminate the indirect discrimination that brings about such underrepresentation.

Assessment of the strength of my arguments should, therefore, proceed independently of the worries mentioned above. In light of the remarks made above, however, another worry might arise. The question would be: if the view that indirect discrimination is not unjust is compatible with its being the case that indirect discrimination ought, morally speaking, to be unlawful, and with measures that are normally taken to counteract its effect, does this article have any significant practical implications at all? I believe the answer is yes, and that this article has two very significant practical implications. The first is that we cannot infer from the fact that a certain group is underrepresented that it is being treated unjustly, just as we cannot infer from the fact that it is overrepresented—witness, my example in the next paragraph—that it enjoys

13. See the discussion in (Schauer 2003), of presumed offenses. For instance, in Bentham's days it was a presumed offence to alter a ship's officially registered name. Obviously, to do so is not an offence in itself, but one can presume that often such a change of name is motivated by a malign reason, i.e. to disguise that the ship has been stolen from its rightful owner.

discrimination in its favour. This is significant, because these inferences of this sort are often made (and often criticized).

The second significant implication is that we will have to think more about what it is that makes cases involving indirect discrimination just or in other ways morally wrong. Take admission rules at Ivy League universities that result in the numerical "overrepresentation" of Asian-Americans. There is an obvious sense in which such rules disadvantage non-Asian-Americans, yet we would not consider this unjust, indirect discrimination. But then why are we inclined to infer this, when the underrepresented group is African-American instead? Enquiries such as the present one force us to try to identify the morally relevant difference. Moreover, they suggest that there are such differences, but that they are not necessarily best thought of in discrimination-related terms. Also, the present enquiry forces us to think hard about the relationship between strict egalitarianism and the injustice of indirect discrimination. In these two ways, and despite the nuances mentioned above, the present article does have significant practical implications. Various forms of affirmative action might well be morally justified, but the present line of argument suggests, surprisingly, that such justification may have little to do with the need to eliminate the injustice of indirect discrimination.

Acknowledgments: Previous versions of this paper were presented at University of Copenhagen, Université catholique de Louvain, the Swedish Congress of Philosophy, and University of Roskilde. I wish to thank Simon Caney, Åsa Carlsson, Axel Gosseries, Rune Klingenberg Hansen, Deborah Hellman, Nils Holtug, Adam Hosein, Magnus Jedenheim-Edling, Sandra Lindgren, Sune Lægaard, Søren Flinch Midtgaard, Thomas Søbirk Petersen, Christian Rostbøll, Ruth Rubio-Marin, Jesper Ryberg, Jens Damgaard Thaysen, and Frej Klem Thomsen for helpful comments. I am also grateful to Larry Alexander for an insightful written response.

REFERENCES

Adarand v. Pena, 515 U.S. 200 (1995), p. 241, http://supreme.justia.com/cases/federal/us/515/200/case.html.

Larry Alexander, "What Makes Wrongful Discrimination Wrongful? Biases, Preferences, Stereotypes, and Proxies" *University of Pennsylvania Law Review* 141.1 (1992), 149–219.

——— "Disparate Impact: Fairness or Efficiency?", *San Diego Law Review* 50.1 (forthcoming).

Andrew Altman, "Discrimination," in E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (2011), http://plato.stanford.edu/archives/spr2011/entries/discrimination/ [accessed January 7, 2014].

Elizabeth S. Anderson, *The Imperative of Integration* (Princeton, NJ: Princeton University Press, 2010).

Richard Arneson, "Against 'Complex' Equality", in David Miller and Michael Walzer (eds.), *Pluralism, Justice, and Equality* (Oxford: Oxford University Press, 1995), pp. 226–252.

William G. Bowen and Derek Bok, *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions* (Princeton, NJ: Princeton University Press, 1998).

John Broome, *Weighing Goods* (Oxford: Basil Blackwell 1991).

Matt Cavanagh, *Against Equality of Opportunity* (Oxford: Oxford University Press, 2003).

G. A. Cohen, *Why Not Socialism?* (NJ: Princeton University Press, 2009).

——— *On The Currency of Egalitarian Justice and Other Essays in Political Philosophy* (Princeton, NJ: Princeton University Press, 2011).

Michael Connolly, *Discrimination Law* (London: Sweet and Maxwell, 2011).

Ronald Craig, *Systemic Discrimination in Employment and the Promotion of Ethnic Equality* (Leiden: Martinus Nijhoff Publishers, 2007), http://www.equality-online.org.uk/equality_advice/positive_action.html [accessed June 3, 2013].

Kimberlé Crenshaw, "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics," in Anne Phillips (ed.), *Feminism and Politics* (New York: Oxford University Press, 1998), pp. 314–343.

DH v. Czech Republic, (Application no. 57325/00) 47 EHRR 3 (2008).

Oran Doyle, Oran "Direct Discrimination, Indirect Discrimination and Autonomy", Oxford *Journal of Legal Studies* 27.3 (2007), pp. 537–553.

Equality Act 2010, Section 19 (1-3), p. 10, http://www.legislation.gov.uk/ukpga/2010/15/contents [accessed June 3, 2013].

Sandra Fredman *Discrimination Law* 2. ed. (Oxford: Oxford University Press, 2011).

John Gardner, "Discrimination as Injustice", *Oxford Journal of Legal Studies* 16.3 (1996), pp. 353–367

Griggs v. Duke Power 401 U.S. 424 (1971), http://www.law.cornell.edu/supct/html/historics/USSC_CR_0401_0424_ZS.html.

Lena Halldenius, "Dissecting 'Discrimination'", *Cambridge Quarterly of Healthcare Ethics* 14.4 (2005), 455–463.

Patricia Hartnett, "Nature or Nurture, Lifestyle or Fate: Employment Discrimination Against Obese Workers", *Rutgers Law Journal* 24.3 (1992—1993), pp. 807–845.

Deborah Hellman, *When is Discrimination Wrong?* (Cambridge: Harvard University Press, 2008).

Nils Holtug, *Persons, Interests, and Justice* (Oxford: Oxford University Press, 2011), pp. 181–201.

Nils Holtug and Kasper Lippert-Rasmussen, "Introduction", in Holtug and Lippert-Rasmussen (eds.) *Egalitarianism: New Essays on the Nature and Value of Equality* (Oxford: Oxford University Press, 2007), pp. 6-7.

Frances Kamm, *Intricate Ethics* (Oxford: Oxford University Press, 2007), pp. 24, 170-173, on the Principle of Secondary Permissibility.

Klemens Kappel and Peter Sandøe, "QALYs, Age and Fairness", *Bioethics* 6.4 (1992), pp. 297–316.

Kasper Lippert-Rasmussen, "The Insignificance of the Distinction between Telic and Deontic Egalitarianism" in Holtug and Lippert-Rasmussen (eds.) *Egalitarianism: New Essays on the Nature and Value of Equality* (Oxford: Oxford University Press, 2007), pp.101-124.

——— *Born Free and Equal? A Philosophical Inquiry Into the Nature of Discrimination* (New York: Oxford University Press, 2013), pp. 38-40.

——— "Distributive Justice and Discrimination", in Serena Olsaretti (ed.), *Oxford Handbook to Distributive Justice* (Oxford: Oxford University Press, forthcoming).

Andrew Mason, "Egalitarianism and the Levelling Down Objection", *Analysis* 61.3 (2001), pp. 246–254.

James W. Nickel, "Discrimination", in Edward Craig and Edward Craig (eds.) *Concise Routledge Encyclopedia of Philosophy* (London: Routledge, 2000), p. 214.

Derek Parfit, "Equality and Priority," in Andrew Mason (ed.) *Ideals of Equality* (Oxford: Blackwell Publishers, 1998), pp. 1 – 20.

Ingmar Persson, "Why Levelling Down Could be Worse for Prioritarianism than for Egalitarianism", *Ethical Theory and Moral Practice* 11.3 (2008), pp. 295–303.

John Rawls, *A Theory of Justice* (Oxford: Oxford University Press, 1971).

Ricci v. DeStefano (Nos. 07–1428 and 08–328) 530 F. 3d 87 (2009), http://www.law.cornell.edu/supremecourt/text/07-1428.

Frederic Schauer, *Profiles, Probabilities, and Stereotypes*, (Cambridge, MA: Harvard University Press, 2003), pp. 224–250.

Michael Selmi, "Was Disparate Impact Theory a Mistake?", *UCLA Law Review* 53 (2006), pp. 701–782.

Shanagan v. UK (Application no. 37715/97) 2001; Council Directive 2000/43/EC of 29 June 2000, article 2(b), http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000L0043:en:HTML [accessed January 5, 2014].

Peter Singer, "Is Racial Discrimination Arbitrary?", *Philosophia* 8.2–3 (1978), 185–203.

Charles A. Sullivan, "The World Turned Upside Down? Disparate Impact Claims by White Males", *Northwestern University Law Review* 98.4 (2004), 1505-1565.

David A. Strauss, "Affirmative Action and the Public Interest", *The Supreme Court Review*, 1995 (1995), pp. 1–43.

Larry S. Temkin, *Inequality* (Oxford: Clarendon Press, 1993).

Alex Voorhoeve and Michael Otsuka, "Why it Matters that Some are Worse Off than Others: An Argument against the Priority View", *Philosophy & Public Affairs*, 37.2 (2009), pp. 171–199.

Michael Walzer, *Spheres of Justice* (Oxford: Basil Blackwell, 1983).

Alan Wertheimer, "Jobs, Qualifications, and Preferences", *Ethics*, 94.1 (1983), 99–112.

Jonathan Wolff and Avner De-Shalit, *Disadvantage* (New York: Oxford University Press, 2007).

Comment on "Associative Duties and the Ethics of Killing in War"

JEFF MCMAHAN

University of Oxford

Seth Lazar's "Associative Duties and the Ethics of Killing in War" (*Journal of Practical Ethics,* Volume 1, Number 1) is an original, rich, challenging, and intricately argued contribution to our understanding of the ethics of war. Its main aim is to explain how fighting in a war can be permissible when warfare inevitably involves the killing of people who are not liable to be killed—a problem that is more extensive than it may seem.

Civilians are almost inevitably killed in war and Lazar accepts that few if any civilians are liable to be killed in war. In principle, of course, a war could be fought without killing civilians, and certainly without killing them intentionally. Yet it is scarcely possible to fight a war, or at least a war in the familiar sense, without intending to kill enemy combatants. Lazar believes, however, that many combatants are not liable to be killed. As I and others have argued, those who fight for a just cause in a just war ("just combatants"), and by permissible means, do nothing to make themselves morally liable to be killed—that is, they do nothing to forfeit or lose their right not to be killed. And Lazar himself has argued, in previous work, that many combatants who fight in wars that lack a just cause ("unjust combatants") are also not liable to be killed. This is because the harm that, as individuals, they threaten to cause is insufficiently great or because the degree to which they are responsible for the harm they threaten is too low. I think he is right about this, though I think the proportion of unjust combatants who are not liable to be killed is in most cases lower than he thinks it is.

The fact that many combatants are not liable to be killed poses a problem for just war theory because, Lazar claims, "contemporary philosophers of the ethics of

war ... until recently unanimously affirmed ... that in justified wars those whom we kill and maim are liable to be killed." Since no contemporary philosopher writing on the ethics of war claims that all the civilians killed as a side effect in justified wars are liable to be killed, Lazar must mean that philosophers have claimed that wars are justified only when those killed *intentionally*, as a means of achieving the war's aims, are liable to be killed. But if this were correct and yet many combatants on both sides were not liable to be killed, there could in practice be no justified wars. Pacifism would be the correct account of the morality of war. To avoid being committed to pacifism, Lazar thinks we should accept that "warfare necessarily involves violating rights," including the rights of those who are killed without being liable to be killed, but also accept that "weightier reasons can override those rights violations, rendering warfare all things considered justified, though unjust." (4)

The main aim of Lazar's argument is to show that associative duties to those to whom we are specially related can sometimes override the rights of people not to be killed, "thus rendering some acts of killing [in war] all things considered justified." (5)

(A brief parenthetical comment. When Lazar writes that philosophers have claimed that a justified war is one in which those killed intentionally are liable to be killed, he also comments that "Jeff McMahan...accepted this commitment without question." (4) Just for the record, I did not accept this uncritically because I did not accept it at all. In a paper published in 2005, for example, I wrote that "while all just wars are morally justified, it seems that not all morally justified wars are just wars. ... [T]here seem to be wars that are morally justified despite their requiring the targeting of those who are innocent in the relevant sense... The form of justification in these latter cases is familiar: in rare circumstances, considerations of consequences override constraints on action that would otherwise be decisive." ["Just Cause for War" *Ethics and International Affairs*, p. 16.])

Lazar develops a plausible non-teleological or non-instrumental account of the significance of special relations and the way in which they ground associative duties. A substantial portion of his essay then seeks to show that associative duties can ground a permission, in a restricted range of conditions, to defend a person to whom one is specially related even at the cost of killing another person who is not liable to be killed. He then argues that this permission can apply in war and can explain the permissibility, at least in certain conditions, of killing combatants who are not liable to be killed. This Associativist Account of the permissibility of certain killings in war provides, he suggests, the best way of avoiding being committed to pacifism.

Its appeal to associative duties supports the claim that "combatants on both sides of a war can, in some cases, fight justifiably." (42) In particular, it explains how unjust combatants "can sometimes be justified in fighting, and on much the same grounds" as those that justify the belligerent action of just combatants. (42)

Lazar contends that, "at least in 1:1 cases, the duty to protect" a person to whom one is specially related "can override the general negative duty not to foreseeably... kill a non-liable person." (19) In other words, it can be permissible for a third party to save the life of an innocent person to whom he is specially related even when, in doing so, he will knowingly kill an innocent or nonliable bystander as a side effect. I will refer to this as Lazar's *central claim*.

Lazar defends his central claim by presenting a transitivity argument. He offers both a stronger and a weaker version of the argument and says that he favors the stronger version. But the stronger version assumes that it is morally *required* to save 5 when doing so would kill an innocent bystander as a side effect of the redirection of a preexisting threat. This is too controversial to be a reliable foundation for his central claim. I will therefore briefly summarize and comment on the weaker version. But my main objection applies equally to both versions.

Lazar claims:

1. It is permissible to redirect a meteor (or trolley) as a means of saving five nonliable people even though this has the foreseen side effect of killing one nonliable bystander.

2. It is permissible to save one to whom one is specially related rather than save five nonliable people.

3. From these two claims he infers that:

4. It is permissible to redirect a meteor (or trolley) as a means of saving one to whom one is specially related even though this has the foreseen side effect of killing one nonliable bystander.

The logic of the argument seems to be this. The moral weight of saving five people is the same in cases 1 and 2. If not killing a nonliable bystander as a side effect has *less* moral weight than saving five, while saving one to whom one is specially related has

greater moral weight than saving five, it follows that saving one to whom one is specially related has greater moral weight than not killing a nonliable bystander as a side effect.

There are various, though related, reasons for doubting the validity of this argument. There are, for example, counterexamples to the claim that "permissible when the alternative is" is a transitive relation. For whether an act is permissible can depend on what the alternatives are. Lazar cites one such counterexample from the work of Frances Kamm. There are others. (See, for example, Derek Parfit, "Future Generations: Further Problems," *Philosophy and Public Affairs* 11 (1982), p. 131.) He argues that the reason why the "transitivity of permissions" fails in Kamm's case does not apply to his argument. He may well be right about that and it may also be true that his argument differs in relevant respects from other similar arguments in which transitivity fails. There are, however, two other concerns that I will merely mention but not pursue.

One is that there may be what Kamm calls "contextual interaction" among the factors in the different cases. It may be, for example, that certain relevant considerations arise in choices in which one option involves killing that do not arise in choices among options that involve only saving and allowing to die.

The second concern is that, despite the assignments of numerical values that Lazar makes for heuristic purposes, the comparisons among killings and lettings die on which the argument depends cannot be precise. This is not because of epistemic limitation but because the relevant values or reasons may in reality be only imprecisely comparable. And when different options are only imprecisely comparable, transitivity may be undermined. (On evaluative imprecision and its significance, see Derek Parfit, "Toward Theory X: Part One" and "Toward Theory X: Part Two," unpublished manuscripts.)

One reason I will not pursue these concerns here is that even if the argument is valid, it requires a further and doubtful assumption to have any serious relevance to the morality of killing in war. The killings in cases 1 and 3 that he claims are permissible are not only merely foreseeable rather than intended but also brought about by the redirection of a preexisting threat—the meteor. This latter fact has often been thought to be part of the explanation of why it is permissible in the standard trolley case to divert the runaway trolley that will otherwise kill five people onto a branch track where it will kill only one person. When Philippa Foot first introduced the trolley case, she contrasted it with a similar case in which to save five patients doctors

must release a gas that will kill one other patient as a side effect, observing that while it seems permissible to kill one as a side effect of saving five in the trolley case, this does not seem permissible in the gas case. ("The Problem of Abortion and the Doctrine of Double Effect," in her *Virtues and Vices,* p. 29.) Commenting on these cases in a later paper, Judith Thomson argued that the reason it is permissible to kill the one in the trolley case but not in the gas case is that in the trolley case one is merely "arranging that something that will do harm anyway shall be better distributed than it would otherwise be." ("The Trolley Case," in her *Rights, Restitution, and Risk,* p. 108.) In the gas case, by contrast, one is creating an entirely new threat to the one.

Since killing in war is almost always done via the creation of a new threat rather than through the redirection of an existing threat, the further assumption that Lazar's argument requires to have significant implications for war is that there is no moral difference between the creation of a threat and the redirection of an existing threat. But, as Thomson contended, this is a dubious assumption.

If in Lazar's first case the only way to prevent the meteor from landing on the five were to blow it up by detonating a bomb that would itself kill a nonliable bystander as a side effect, it seems that this would not be permissible. And it seems the same would be true if in his third case the only way to prevent the meteor from killing one's child were to create an explosion that would kill a nonliable bystander. This should not be surprising given the difference between people's intuitive reaction to the trolley case and their reaction to the gas case. But if it is right that there is a moral difference between killing via the redirection of a preexisting threat and killing via the creation of a threat, it seems that the plausibility of Lazar's transitivity argument is restricted to cases in which the killing is done via redirection. This means that the scope of the argument is highly limited and that it is virtually irrelevant to the justification of killing in war, which is seldom done by the redirection of a preexisting threat.

Some people, of course, believe that it is permissible for an individual to act in self-defense or self-preservation in a way that will create a threat that will kill a nonliable bystander as a side effect. But this is a minority view. Even those who claim that there is an *agent-relative permission* to kill a wholly innocent or even nonresponsible person who *threatens* one's life are usually reluctant to accept that this permission extends to an act of self-preservation that creates a threat that will kill a bystander as a side effect.

But even it if is not permissible to kill a nonliable bystander as a side effect of an act of self-preservation, it might be permissible to kill such a person as a side effect

of an act of saving a person to whom one is specially related. For an agent-relative permission and an associative duty are distinct sources of reasons, and it may well be that at least some of one's special relations to others are sources of stronger reasons than is the relation of identity one bears to oneself. It might, for example, be permissible for a parent to kill a nonliable bystander as a side effect of saving her child even when it would not be permissible for the child to save himself if doing so would unavoidably kill the same bystander as a side effect.

But even a view of this sort has little relevance to the permissibility of killing in war. One reason is that such a view is plausible, if at all, only in the case of the most significant special relations, such as the relation between a parent and child. It would not be permissible to kill a nonliable bystander as a side effect of saving one's neighbor, even though being neighbors is a special relation that has a certain degree of moral significance. But in war a soldier almost never knows that an act that would kill an enemy combatant is necessary to save the life of someone as closely related to him as his child. In general, the most he can know is that the act may slightly reduce the risk to someone closely related to him of being killed by the enemy.

There is, however, one exception to this, which is that a combatant can sometimes know that his killing an enemy combatant is necessary to save the life of one of his comrades-in-arms. Lazar cites the claim of J. Glenn Gray and others that the most important factor that motivates combatants to continue to fight rather than to flee or surrender is the compulsion they feel to protect their close comrades. A specialist on the psychology of war reports that "in military writings on unit cohesion, one consistently finds the assertion that the bonds that combat soldiers form with one another are stronger than the bonds most men have with their wives." (Quoted in Lt. Col. Dave Grossman, *On Killing,* 1995, p. 149.) Perhaps this special relation strengthens the justification that combatants have for killing enemy combatants even when the latter may not be liable to be killed.

But I doubt that this is true. Suppose the combatants under attack are just combatants and the enemy combatant that one of them must kill to save the other is an unjust combatant. In that case, the unjust combatant threatens a just combatant with death and hence is morally liable to be killed, even if the degree of his responsibility for the threat he poses is low. There is already a liability justification for killing him; hence the appeal to an associative duty is otiose.

If instead the combatants under attack are unjust combatants and the enemy combatant who threatens one of them is a just combatant, it is then very unlikely

that the special relation between the two unjust combatants could justify the killing of the just combatant. There are at least four reasons why this is so, some of which are recognized by Lazar himself.

First, the bonding that the unjust combatants have experienced is not so much a *reason* for fighting as it is a *result* of fighting. The bond has developed because of their having been in combat together. But because their war is unjust, the bond has arisen because of their shared participation in an activity that is objectively wrong. Indeed, they did wrong to get themselves into the situation in which the bond developed and now motivates them to kill people who are not liable to be killed. As Lazar acknowledges, following Thomas Hurka, the contaminated nature of the relation between them diminishes or even vitiates altogether its moral significance.

Second, as Lazar also concedes, the associative duty to protect someone to whom one is specially related has at most only a weak application when the that person is liable to be harmed. And when an unjust combatant will otherwise kill a just combatant, that unjust combatant is liable to be killed. Thus, even if the special relation that the one unjust combatant bears to the other were highly morally significant, it would not justify a combatant in killing a nonliable person as a means of defending his comrade-in-arms against an attack to which he was liable.

Third, Lazar's central claim concerns the permissibility of killing a nonliable bystander. But a just combatant is not a bystander; he is a just threatener vis-à-vis the unjust combatant he threatens. Killing him is therefore wrong for two distinct reasons: it would not only wrong him but also prevent him from achieving his just aim.

Fourth, and finally, Lazar's central claim concerns the justifiability of killing a nonliable person *as a side effect*—that is, an unintended effect. But the killing of a just combatant as a means of saving one's fellow unjust combatant is an intended killing, and the constraint against killing a nonliable person as a means is, as Lazar recognizes, stronger than that against the killing of a nonliable person as a side effect.

This is perhaps the main reason why Lazar's central claim has little relevance to the justification for killing in war. That claim is that it can be permissible to kill a nonliable person as a side effect of the redirection of a preexisting threat away from someone to whom one is significantly specially related. But the killing of combatants in war, which is what needs to be justified in principle if pacifism is to be avoided, is much more often intended than unintended and is almost always accomplished by the creation rather than the redirection of a threat.

It is, however, unclear what Lazar takes the scope of his argument to be. While his central claim is about the justifiability of killing as a side effect, he is clearly aware that the refutation of pacifism requires a justification for at least some intentional killing of combatants, and particularly unjust combatants, who are not liable to be killed. Thus, he writes that "the real challenge is to show that [associative duties] can license some intentional killing of combatants, without also permitting intentional killing of noncombatants." (34)

His effort to meet this challenge begins with four reasons why it is more seriously wrong *intentionally* to kill nonliable noncombatants than it is to kill nonliable combatants. Of these four reasons, the one to which he devotes most space is that "it is more wrongful to kill nonliable people who are defenceless and vulnerable than to kill those who can fight back, or who are less vulnerable." (36) Although common, this claim has always seemed implausible to me. It implies, for example, that it is more seriously wrong to kill just combatants using bombers or long-range artillery than to kill them in close combat—so that an unjust combatant could say in mitigation, "I killed him, which was wrong, but at least I didn't kill him from a safe distance." But I will not discuss this further here, as the more important question is whether there are reasons to think that combatants' associative duties can make it permissible for them intentionally to kill other nonliable combatants when the killing would be impermissible in the absence of the associative duties. (For a powerful critique of the idea that it is more seriously wrong to harm nonliable people who are defenseless than to harm otherwise similar people who are not entirely defenseless, see Jonathan Parry, "Community, Liability, and Just Conduct in War," forthcoming.)

Lazar's discussion of this issue is confusing because of the puzzling way in which he uses the terms "eliminative" and "opportunistic" to apply to acts of harming or killing. On page 33, for example, he restates his central claim in this way: "our duties to protect those we share valuable relationships with can override the duty not to kill a nonliable person, at least in 1:1 cases where the victim is killed foreseeably and eliminatively, rather than intentionally and opportunistically." In discussions of the ethics of harming and killing, the term "foreseeable" is often used as shorthand for "foreseeable but unintended" and it is reasonable to assume that that is what Lazar means here. But if a person is killed only foreseeably and not intentionally, he is *not* killed eliminatively, in the sense in which the latter term is used in the literature. For the distinction between eliminative and opportunistic killing has hitherto been understood as a distinction between two forms of *intended* killing.

Warren Quinn, who introduced the distinction, first defined *direct* harmful agency as "agency in which harm comes to some victims...from the agent's deliberately involving them in something in order to further his purpose precisely by way of their being so involved (agency in which they figure as *intentional objects*)" and *indirect* harmful agency as "harmful agency in which either nothing is in that way intended for the victims or what is so intended does not contribute to their harm." He then suggests that the revised doctrine of Double Effect that he proposes might "strongly discriminate... against direct agency that benefits from the presence of the victim (direct *opportunistic* agency) and more weakly discriminate...against direct agency that aims to remove an obstacle or difficulty that the victim presents (direct *eliminative* agency)." ("Actions, Intentions, and Consequences: the Doctrine of Double Effect," *Philosophy and Public Affairs* 18 (1989) , p. 344.) Both eliminative and opportunistic agency are thus defined as forms of "direct" agency—agency that affects a victim intentionally. One reason for understanding the distinction this way is that harm that is merely a foreseen side effect is not instrumental in eliminating a threat. But to call harming or killing eliminative is to acknowledge that it is instrumental in eliminating a threat from the person harmed or killed rather being a means of avoiding a threat of which that person is not the cause.

Lazar, by contrast, defines eliminative killing as killing in which "the killer derives no benefit from the victim's death that he would not have enjoyed in the victim's absence." (19) He then observes that in his case 3 when the agent diverts the meteor away from the person to whom he is specially related but toward a nonliable bystander, the killing of the bystander is eliminative. This is not the way the term has been used by others, but this would not matter if it were not that the deviant use makes it unclear what Lazar means to say about the role that associative duties might have in justifying the most common form of killing of combatants in war—namely, killing to eliminate the threat that a combatant poses. Thus, in restating his central claim in the essay's penultimate paragraph, he writes that "the argument advanced above was that in 1:1 cases, A's duty to protect B from lethal harm can justify the foreseeable infliction of eliminative lethal harm on C, but that it cannot justify opportunistically harming C." (43) The problem with this statement is that the lethal harm that Lazar describes as both foreseeable and eliminative could be either unintended or intended. The natural interpretation would be to read "foreseeable" as implying unintended. Yet the context is a discussion that purports to explain how the appeal to associative duties can help to justify the *intended* killing of combatants in war. And

it is the justification of intended killing that is needed to support Lazar's claim that unjust combatants "can sometimes be justified in fighting, and on much the same grounds as" just combatants. Yet what he concedes, in the same sentence, is that his argument cannot justify is *opportunistic* harming, not that it cannot justify *intentional* harming. Here he is repeating his earlier concession that associative duties "cannot justify the opportunistic killing that war inevitably involves." (34) In these passages I find that I simply do not know what he means to be asserting. The suggestion that war inevitably involves opportunistic killing suggests that he may understand all killing as a means to be opportunistic. In that case, the killing of enemy combatants as a means of averting the threat they pose would be opportunistic. But this understanding of the term is incompatible with the definition he gives on page 19, which is consistent with Quinn's. Yet according to this definition of opportunistic killing, it is not true that war inevitably involves opportunistic killing. The aims that are served by the killing of enemy combatants in war rarely require the presence of those combatants for use as a means. If the combatants were not there, there would be no need to fight. What war inevitably involves—indeed almost necessarily involves—is killing of enemy combatants that is eliminative, in Quinn's sense, rather than opportunistic (though sometimes killing that is primarily eliminative can have an opportunistic dimension as well, as when killing is intended not only to eliminate the threat the immediate victims pose but also to intimidate the victims' fellow combatants).

In the end, I cannot find an argument in the essay for the extension of Lazar's central claim so that it also applies to the intentional, eliminative killing of nonliable unjust combatants. This is not to say that I think such an argument cannot be made. I suspect that it can. That is, I do not find it implausible to suppose that a just combatant's associative duties do in some cases strengthen the justification for the eliminative killing of unjust combatants. But Lazar's confusing and, I think, unnecessary invocation of the distinction between eliminative and opportunistic killing has been an impediment to his ability to produce such an argument.

There are, moreover, other good reasons why it can sometimes be permissible to kill nonliable unjust combatants; so we need not fear being compelled to embrace pacifism. One such reason is that unjust combatants are themselves responsible for making it reasonable for just combatants to believe that they are liable to be killed. Another is that when they are killed intentionally because their adversaries cannot know that they are not liable, killing them is morally like killing as a side effect. It does not come within the scope of the constraint against the intentional killing of

people who are not liable to be killed. (For defenses of these claims, see Jeff McMahan, "Who is Morally Liable to be Killed in War," *Analysis* 71 (2011), pp. 555-59.) A third and perhaps more important reason is that even when unjust combatants are not liable to be killed, they are normally liable to some degree of harm less than that involved in being killed. It might therefore be permissible to kill them if part of the harm they would suffer in being killed could be justified on the ground that they are liable to it, while the remainder could be justified on ground that it is the lesser evil when the alternative is to allow the achievement of their unjust cause.

Finally, I also cannot find an argument in the essay that explains how an appeal to associative duties can help to vindicate Lazar's conclusion that "combatants on both sides of a war can, in some cases, fight justifiably." That is, I cannot find an argument that shows that the associative duties of unjust combatants can justify their intentional killing of just combatants. And here I do think that no such argument, or at least a plausible one, can be made.

A Reply to McMahan

SETH LAZAR

Australian National University

Jeff McMahan raises some valid concerns about the paper, and I am grateful both for his taking the time to do so, and for the opportunity to clarify my central argument. Rather than attempt a blow-by-blow defence, I'll just concentrate on three major points, before summing up.

I elliptically misattributed to McMahan the Walzerian view that intentional killing in war is justified only when the target is liable to be killed. However, throughout McMahan's work (until recently) the role of lesser evil justification has been radically circumscribed, confined to situations where intentionally killing the nonliable is necessary to avert an unusual catastrophe, of the sort that justified war does not typically aim to avert (of course Walzer too thought intentionally killing the nonliable permissible in supreme emergencies). Throughout his oeuvre, McMahan has argued that most combatants on the unjust side are sufficiently responsible for unjustified threats to be liable to be killed. I intended to target this thesis. In ordinary wars (which are not fought to avert supreme emergencies) many combatants who are intentionally killed are not liable to be killed, so if killing is justified, it is as a lesser evil. Since we do think that wars can sometimes be justified, even if not to avert an unusual catastrophe, this suggests that lesser evil justification must play a greater role in just war theory than Walzer, and until recently McMahan have thought. McMahan's recent move towards a similar view is a welcome evolution, but an evolution nonetheless.

McMahan's central objection is that my discussion lacks an explanation of just how associative duties can override the duty not to intentionally kill a nonliable person—since my central case, used to show that associative duties can override serious general negative duties, does not involve intentional killing. Again the shortcoming is one of clarity of exposition on my part. To clarify, the argument is this: not all killings of nonliable people are equally wrongful. These killings differ in

their qualitative, agential component. My examples show (so I claim) that associative duties can justify at least some kinds of killing—this is in stark contrast to the standard philosophical view, that associative duties cannot justify overriding any serious general negative duties. I then argue that some of the killing that warfare involves is no more wrongful than this. I focus, in particular, on showing that killing combatants is less wrongful than killing noncombatants, because of considerations such as vulnerability, recklessness, consent and so on. I concede that my focus was too much on showing that killing noncombatants is especially wrongful—but just the same arguments work to show that intentionally killing nonliable combatants is relatively less wrongful. McMahan's own view is that there is a step-change between intentional killing, of any description, and any sort of unintended but merely foreseen killing. I find this implausible. There is simply a range of different agential factors, which render killing more or less wrongful. Intention is one of them, but it is not fundamentally distinct from the others that I describe—in particular whether one's agency is eliminative or opportunistic, whether one's actions are reckless, whether the victim is vulnerable etc. On the specific definition of eliminative agency that I use, it does indeed depart from Quinn's notion that eliminative harmful agency must be intentional, but I think my interpretation is an improvement on Quinn's—I defend it at length elsewhere.

Some of McMahan's other worries I think are adequately foreshadowed and addressed in the text—the worry that soldiers only rarely defend those with whom they themselves share valuable relationships is discussed at length on pp.30-33; the worry that their relationships with their fellow combatants might be tainted by their contributing to an unjust war is addressed on pp.40-41; and I stress on p.41 that on the high threshold view of liability that I affirm, even those on the overall unjust side will often not be liable, so their associates will often have duties to defend them.

Ultimately the argument of the paper is simple—perhaps simpler than I made out! At least some combatants on the unjust side are fighting to protect nonliable combatants and noncombatants on their side, both performing associative duties that they owe directly to those whom they protect, and acting on behalf of the community at large to perform the duties that its members owe to protect those with whom they have valuable relationship. Associative duties to protect those with whom we share valuable relationships can justify some of the less wrongful forms of killing of a nonliable person. Killing nonliable combatants in war (including those on the just side, note) is one of the less wrongful forms of killing a nonliable person. Killing nonliable

noncombatants is one of the more wrongful such forms. So these associative duties can justify killing combatants in war (even on the just side), without also justifying killing noncombatants.