

Journal of Practical Ethics

❧ Volume 1, Number 1. June 2013 ❧

CONTENTS



Introducing the Journal of Practical Ethics <i>Roger Crisp & Julian Savulescu</i>	I
Associative Duties and the Ethics of Killing in War <i>Seth Lazar</i>	3
Biotechnology, Justice and Health <i>Ruth Faden & Madison Powers</i>	49
Situationism and agency <i>Alfred R. Mele & Joshua Shepherd</i>	62

Editors in Chief:

Roger Crisp (Oxford University)

Julian Savulescu (Oxford University)

Managing Editor:

Bennett Foddy (Oxford University)

Associate Editors:

Tom Douglas (Oxford University)

Guy Kahane (Oxford University)

Kei Hiruta (Oxford University)

Editorial Advisory Board:

John Broome, Allen Buchanan, Tony Coady, Ryuichi Ida, Frances Kamm, Philip Pettit

Editorial Assistant:

Miriam Wood

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

The material in this journal is distributed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported licence. The full text of the licence is available at:

<http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode>

© University of Oxford 2013 except as otherwise explicitly specified.

ISSN: 2051-655X

Introducing the Journal of Practical Ethics

ROGER CRISP

Oxford Uehiro Centre for Practical Ethics, University of Oxford

JULIAN SAVULESCU

Oxford Uehiro Centre for Practical Ethics, University of Oxford

The *Journal of Practical Ethics* is a new open access, interdisciplinary journal in applied moral philosophy and related areas of philosophy, including political and legal philosophy. It is supported by the Uehiro Foundation on Ethics and Education and we are most grateful to the Foundation for making the journal possible. Areas to be covered will include medical ethics, environmental ethics, business ethics, professional ethics, the ethics of personal life and others.

There are already many outstanding academic journals covering issues in practical ethics, including *Ethics*, *Philosophy and Public Affairs*, *The Journal of Applied Philosophy*, *Bioethics*, *Utilitas*, and *Ethical Theory and Moral Practice*. Why have we launched this new journal in what might seem an already crowded market-place?

We believe that the ideas and arguments of many moral and political philosophers are of significant relevance to problems in contemporary life. Not only are these arguments of interest to the general public, but they are of relevance to political and social leaders, legislators and civil servants. However, there is less than optimal penetration of this philosophical work beyond the confines of academe. This is partly because the leading journals are not open access. But it is also because the articles in these journals tend to be written for other academics, who are assumed to be familiar with the philosophical background. Our journal is not just for professional academics and their students, though we hope they will find much to interest them in our pages. We aim to disseminate excellent research in practical and applied ethics to a

broad, global audience including non-academic as well as academic readers, with real and significant impact.

The journal will differ from many already in the field in that it will extend beyond pure research output to include an educational, public access component. Therefore not all articles will be entirely original research, although many will be. Articles may consist of the application of more theoretical work to a practical issue, or in a restatement of an author's position in a more reader-friendly format. We aim to convey the ideas and ways of thinking of the best philosophers to a worldwide audience and to show how they can be helpfully applied to a practical issue. Articles will aim for a level of accessibility to a non-specialist audience commensurate with, for example, the level of philosophy articles in *The Boston Review*, The Stone column in the *New York Times Online* or *Prospect*. The journal will seek to avoid technical jargon and obscure terminology. The hope is that anyone with a serious interest in the issue at stake, anywhere in the world and beyond the walls of any university, will be enabled to read work by outstanding philosophical thinkers, such as those in our first issue, just by opening the internet, and to acquire insight and food for thought not easily available elsewhere. We will adopt an active approach to promoting our content on the web. We shall use social media and any other resources available to us to increase awareness of our journal and traffic to our articles.

As is standard in intellectual magazines and papers aiming at a broad readership, all our articles will be commissioned and an honorarium will be offered to each author. We shall ensure as broad a representation of views as possible, subject always to strict requirements of rigour and clarity of argumentation. The journal will not adopt any specific ideological, moral, or political position, and articles will be peer-reviewed. Though the journal will be published in open access form online, print copies will be available at cost. Both online and print versions will appear as familiar print journals with full consecutive pagination.

We hope that many thousands of new readers from around the world will benefit from our pages over the coming years, and that our journal will make a genuine and sustained contribution to thought about some of the deep problems facing humanity in the twenty-first century, and to action needed to address these problems.

Associative Duties and the Ethics of Killing in War

SETH LAZAR

Department of Philosophy, Australian National University

ABSTRACT

This paper advances a novel account of part of what justifies killing in war, grounded in the duties we owe to our loved ones to protect them from the severe harms with which war threatens them. It discusses the foundations of associative duties, then identifies the sorts of relationships, and the specific duties that they ground, which can be relevant to the ethics of war. It explains how those associative duties can justify killing in theory—in particular how they can justify overriding the rights to life of some of those who must be killed to win a war. It then shows how these duties can be operationalised in practice: first, showing how soldiers who fight on behalf of their community can act on reasons that apply to the members of that community; second, showing that the argument from associative duties does not prove too much—in particular, that it does not license the intentional killing of noncombatants in war.

INTRODUCTION

Many of us believe that pacifism is mistaken. Thus, warfare, composed though it is of killing and maiming, can sometimes be justified. We also believe humans enjoy fundamental moral protections against being killed and maimed—commonly expressed in the language of human rights. The challenge is to render these two commitments mutually consistent. There are two possibilities: first, we could argue that in justified wars those whom we kill and maim are liable to be killed, that is, they do not

enjoy the protection of their rights—they are lost, or forfeit, so killing and maiming is just, because rights-consistent. Second, we could concede that warfare necessarily involves violating rights, but argue that weightier reasons can override those rights violations, rendering warfare all things considered justified, though unjust.

Contemporary philosophers of the ethics of war, despite their other profound disagreements, until recently unanimously affirmed the first option. Michael Walzer started the trend, arguing that ‘a legitimate act of war is one that does not violate the rights of the people against whom it is directed’ (Walzer, 2006, p. 135). Jeff McMahan, for all his other criticisms of Walzer, accepted this commitment without question (McMahan, 1994, 2004, 2009; although see McMahan, forthcoming). Likewise Walzer’s supporters (Benbaji, 2008; Emerton & Handfield, 2009; Steinhoff, 2008), and his many opponents (Altman & Wellman, 2008; Coady, 2008; Fabre, 2009; Kamm, 2004; McPherson, 2004; Miller, 2007; Øverland, 2006, 455-475; for dissenting views see Kutz 2008 and Shue 2008). The dispute between these camps is over who loses their rights in war: Walzerians advocate the ‘symmetrist’ position that combatants on both sides of any war lose their rights to life, while noncombatants on both sides retain them; anti-Walzerian ‘asymmetrists’ think that combatants and perhaps some noncombatants whose side went to war unjustifiably—hereafter combatants-U and noncombatants-U—lose their rights against attack, but combatants and noncombatants on the justified side—combatants-J and noncombatants-J—retain them.

I think this uncritical endorsement of Walzer’s starting assumption is a mistake—at least if we want to justify a plausible set of armed conflicts, while affirming a plausible theory of the right to life (see Lazar, 2009a and Lazar, 2010a; for similar views see Kutz, 2008 and Shue, 2010). If all combatants-U must be liable to be killed for warfare to be permissible, then we have to set the threshold of responsibility for liability to be killed low enough to ensure that even the most inept, inactive combatant is sufficiently responsible to bear liability. But if we endorse this low threshold view, then at least in modern states, we will render whole adult populations liable to be killed, since we are all responsible, if only minimally, for our states’ warfighting capacities—through our tax-paying, voting, contribution to war-related industries, and so on. If mere contribution or innocent contribution were sufficient for liability to be killed, very few would escape the liability net. Most will find this untenable, so must argue that liability presupposes something thicker: there must be a ‘fit’ between the degree of responsibility or contribution, and the resulting fate. But once we endorse this

high threshold view, we find that many of the combatants-U whom we need to kill are not sufficiently responsible, or do not contribute enough, to be liable to be killed.

For those who reject collective liability—which includes all the just war theorists mentioned so far—liability can be grounded only in the assessment of each potentially liable individual's behaviour. As his critics argued, Walzer's generalisation over all combatants is unacceptable—but so is his critics' generalisation over all combatants-U. Considered as individuals, at least on the more plausible high threshold view of liability, some combatants-J are liable to be killed; not all combatants-U are liable. If individual rights are inviolable, we must ensure that each individual whom we kill or maim is liable to that fate. Yet warfare is too messy for this demand to be feasible. Most killing happens from a distance, and the usual targets are coordinates rather than individuals. The relevant actors frequently cannot discriminate between those that are and are not liable to be killed. Any theory of our right to life that is sufficiently indiscriminate to render this carnage consistent with that right, is surely not discriminating enough to be a plausible theory of our right to life. In all likely wars, we will intentionally and unintentionally kill and maim many nonliable people. Wars cannot be wholly just. If they must be entirely just to be justified, we should endorse pacifism.

This paper proposes one strand in an alternative solution—one that affirms the high-threshold view of liability to be killed, and therefore concedes that the rights-respecting war is an unattainable ideal, but maintains that warfare can nonetheless sometimes be justified. There are as many different approaches to this task as there are competing values, whose realisation might sometimes require the use of lethal force. But in this paper I focus on one class of widely neglected reasons. Most of us share a number of morally important relationships with those closest to us—our family, friends, and other loved ones. Combatants enjoy similarly significant relationships with their comrades-in-arms. When aggressors attack, they threaten those with whom we share these relationships—our associates. Sometimes we can protect our associates only if we fight and kill. We have duties to protect our associates, grounded in the value of these special relationships. Our armed forces are the executors of those duties. When they fight, those duties may clash with the rights that they must violate to win the war. In some cases, the associative duties to protect can override those rights, thus rendering some acts of killing all things considered justified. I call this the Associativist Account of (part of) what justifies killing in war.

The motivation for endorsing the Associativist Account is simply that ordinary

understandings of the justification of warfare are characterised by extreme partiality, and yet contemporary just war theory is almost exclusively impartial in orientation. I think it is commonplace to suggest that, when people justify fighting to themselves, their desire to protect their loved ones, their comrades-in-arms, and their community, are at the forefront of their reasoning. In a prominent study of the motivations and mindset of twentieth century armed forces, for example, Joanna Bourke notes that

In a survey of 568 American infantrymen who had seen combat in Sicily and North Africa in 1944, men were asked what was the most important factor enabling them to continue fighting. Leadership and discipline, lack of alternatives, vindictiveness, idealism, and self-preservation ('kill or be killed') were scarcely mentioned. Rather (after simply desiring to 'end the task'), combatants cited solidarity with the group and thoughts of home and loved ones as their main incentives. (Bourke, 1999, p. 142)

In his reflections on his own wartime experiences, philosopher J. Glenn Gray concurs:

The fighter is often sustained solely by the determination not to let down his comrades... Numberless soldiers have died, more or less willingly, not for country or honour or religious faith or for any other abstract good, but because they realized that by fleeing their post and rescuing themselves, they would expose their companions to greater danger. Such loyalty to the group is the essence of fighting morale. (Gray, 1998, p. 40; for further examples from World Wars One and Two, and the recent Iraq conflict, see Barnham, 1975; Lee, 2006, p. 118; Nicol, 2007, pp. 76, 114, 152-3, 214, 254; Remarque, 1996, pp. 117, 145).

Obviously it is possible that this is just the sort of chauvinism that morality should condemn—all the more so given how high the stakes are in war. This paper, however, is motivated by a curiosity as to whether this sort of commonsense partiality has any place in the morality of war.

Notice, though, that the Associativist Account is not a comprehensive theory of what justifies killing in war. The best such theory will include, alongside the arguments discussed here, a theory of permissible killing in self-defence and other-defence, including a defence of the high threshold approach to liability, as well as an

account of the independent moral significance of goods such as political sovereignty and territorial integrity. My goal in advancing the Associativist Account is simply to insist that reasons of partiality are relevant to that more comprehensive theory.

My argument for the Associativist Account will be restricted in one further respect. A full defence should invoke duties that we owe to compatriots *qua* compatriots. I will not argue for these duties here, since doing so is itself a challenging task and the subject of significant debate (though see Lazar, 2010b). Instead, I will argue that the duties that we owe to those with whom we share our deepest, most valuable personal relationships—which often overlap at least with residency in the same territory, if not with co-citizenship—can justify killing nonliable people in war. This constraint allows me to focus on the class of associative duties that are the least contentious, as well as to raise interesting questions about how the reasons that apply severally to the members of a community can be translated into justifications upon which some subset of the community can act.

The paper proceeds as follows. I begin by identifying the relevant class of special relationships, and the associative duties that they ground. I focus on family, friendship, and fellowship of arms. Although I cannot offer a full theory of associative duties in this paper—and wish to remain neutral as to their ultimate justification—I will sketch the main schools of thought regarding how our most valuable relationships ground morally weighty reasons, which can be relevant to the justification of killing in war. In particular, I argue that associative duties can require the duty-bearer to bear greater costs than otherwise comparable general duties—that is, they are more *stringent* than general duties. I then go on to argue that associative duties are also graver than general duties, insofar as they weigh more heavily against other moral considerations. In particular, I argue that our associative duties to protect those with whom we share valuable relationships can sometimes override our general negative duties not to harm others (duties that correspond to the rights with which this introduction began). Here my view defies the philosophical orthodoxy: that associative duties cannot override serious general negative duties (Ashford, 2003; Jeske, 1996; Keller, 2007; Kolodny, 2002; McMahan, 1997; Pogge, 2002; Scanlon, 1998; Scheffler, 2002; Seglow, 2010; in the context of war see Lefkowitz, 2009; for a more sympathetic view, see Kamm, 2004 and Hurka, 2005).

Having shown that when we fight to protect our loved ones, we are justified in imposing more costs on innocents than when we fight to protect strangers, I go on to show how that view can be operationalised in the morality of war. This involves

two tasks: first, showing how soldiers who fight on behalf of their community can justify their actions as the execution of the associative duties that the members of that community owe to protect their loved ones; second, showing that the argument from associative duties can justify the sort of killing that a theory of war's morality particularly needs to justify. The paper then goes on to discuss the restrictions on the Associativist Account; the last section concludes.

GROUNDING ASSOCIATIVE DUTIES

My goal in this paper is to see what follows for the morality of war from accepting that we have weighty duties to protect our associates. I cannot develop a full theory of those duties, while also pursuing that more applied goal. However, I can adumbrate some of the theoretical groundwork. I shall introduce the idea of associative duties, identify plausible grounds for their ultimate justification, and show which relationships ground duties relevant to the morality of war.

I understand duties as a kind of moral reason, distinguished by being non-voluntary, having a particular weight, and retaining their force when overridden. Thus, if you have a duty to ϕ , and no equally weighty moral reason not to ϕ , then you may not permissibly choose not to ϕ . Sometimes, however, you have other reasons not to ϕ , for example a duty to ψ that is impossible with ϕ -ing. If the latter duty is weightier than the former one, you ought to ψ . And yet, though the ϕ duty was overridden, it retains its force: with respect to that duty and its beneficiary, you have acted wrongly, even if ψ -ing was all things considered justified.

Impersonal duties are owed to nobody, interpersonal duties to other person(s) and personal duties to oneself. General duties are interpersonal duties owed to everyone, and special duties are interpersonal duties owed to specific people, because of some interaction or relationship. Associative duties are a subclass of special duties that are owed in virtue of a morally important relationship.

There are two broad schools of thought about the underlying justification of associative duties.¹ The first justifies associative duties teleologically, by appealing to the value of the states of affairs in which they are realised. In its simplest form, the argument is this: acknowledging associative duties is a necessary condition of being in

1. Another faction argues that our relationships ground only non-moral reasons. Nonetheless, insofar as those reasons are relevant to what is all things considered permissible, this difference need not concern us here. See, for example, (Wolf, 1992). Samuel Scheffler argues that these are genuinely moral reasons in (Scheffler, 2010)

a particular kind of deep personal relationship. Since deep personal relationships are either vital constituents of our well-being, or necessary conditions of realising other goods, we ought to acknowledge associative duties (Brighouse & Swift, 2006, p. 95; Cottingham, 1983, pp. 89-90; Cottingham, 1986, p. 369; Etzioni, 2002, p. 596; Friedman, 1991, p. 820; Hardimon, 1994, p. 535; Horton, 2006, p. 437; Keller, 2006, p. 265; Mason 2000, p. 99, McMahan, 1997, pp. 118, 124; Miller, 2005, p. 65; Raz, 1989, pp. 14-5). These bare bones can be fleshed out in several different ways.

The first divide is between those for whom this is an argument within moral theory, and those who view it as a constraint on which moral theories we can plausibly adopt. The first view is more common, but Samuel Scheffler, for example, argues that 'people's interest in obtaining the rewards of special relationships is so strong that morality cannot possibly fail to accommodate it' (Scheffler, 2002, pp. 59, 93).

The second divide is between those who regard associative duties as partly constitutive of these relationships, and those who focus on duties that are instrumental to the relationships' realisation.² Among the former group, some think that the duties constitute these relationships in virtue of patterns of social practice (e.g. Raz 1989) while others are more essentialist, seeing the duties as constitutive of such relationships regardless of context.

The third divide is between those who think these relationships are constitutive of well-being, and those who think they are instrumental to it. On the former view, enjoying deep personal relationships is a crucial component of what makes a life go well (Mason, 2000, pp. 108-9; Horton, 2006, pp. 436-8; 2007, p. 7; Raz, 1989, pp. 20-1). On the second, relationships generate other goods, such as pleasure, which ground their significance (Etzioni, 2002, p. 596; Friedman, 1991, p. 820; Tamir, 1993, p. 97; Scheffler, 2002, pp. 59, 93). One common specification of the latter claim is that deep personal relationships give special access to the needs of some others, enabling us to advance their well-being in ways non-associates cannot (e.g. Goodin, 1985).

This opens up two broad camps within the teleological justification of associative duties: instrumentalists will argue that some relationship is conducive to well-being, and the duty conducive to the relationship; constitutivists will argue that these relationships are constitutive components of well-being, and the duty partly constitutive of the relationship. Hybrid theories will be constitutivist on one point, but instrumentalist on the other.

In contrast with this approach, some have advanced a nonteleological alterna-

2. The former is much the more common view, but for the alternative see (Lazar, 2010b)

tive.³ The view is this: recognising that relationships have properties which ground reasons, including duties, is a matter of responding appropriately to those properties. To fail to respond appropriately to those properties—by breaching an associative duty, for example—is not only to bring about a bad state of affairs, but also to express in one's actions the wrong attitude towards the relationship's valuable properties. Just as a breach of a general duty is a failure to respond appropriately to the victim's status as a member of the moral community, a breach of an associative duty is a betrayal of one's associate as a member of this valuable relationship.

To show that a relationship x grounds associative duties, then, we need to show first that its properties, x^{I-n} , could directly give weighty reasons for action, independently of any other properties to which they conduce. This is similar to saying that they are non-instrumentally valuable. This is, first, because otherwise the reasons grounded in x^{I-n} would not necessarily retain their force when overridden. If x^{I-n} only give reasons insofar as they conduce to y , then if some z better serves y , but at the expense of x , there is no cost in ignoring the reasons given by x^{I-n} . Second, if x^{I-n} do not directly ground moral reasons, then those reasons are properly attributed to the properties that do in fact ground them. If special relationships were only important insofar as they contributed to economic growth, for instance, then any reasons they could give would derive from the importance of economic growth, not from their own properties. Third, since duties are very weighty reasons, both demanding of their bearer, and potentially overriding or excluding other moral reasons, for a relationship to ground duties it would have to be of significant moral importance and thus able to give direct, strong moral reasons.

Second, we need to show that recognising an associative duty counts as the appropriate response to those non-instrumentally valuable properties. As with the teleological account, this can be shown in two ways—either the duty is importantly instrumental to preserving those valuable properties; or it is constitutive of the relationship, such that if you denied it, you would be demonstrating a misunderstanding of the relationship's character and significance.

My own view is that the nonteleological justification for associative duties is much more successful than the teleological alternative, and as such I assume the nonteleological approach throughout the rest of this paper. However, the task of identifying an associative duty on either view is similar: first one must show that the

3. A view like this is endorsed by in (Scheffler, 2000, ch. 6). It differs significantly from the view developed in earlier chapters. It is developed in depth in (Lazar, Unpublished)

relationship has properties that are of significant non-instrumental value (i.e. that give strong, direct reasons for action); then one must show that the duty is either a necessary condition of preserving those properties, or is otherwise constitutive of the relationship, or of an appropriate recognition of the relationship's value. As such, most arguments to show that there are associative duties on the nonteleological approach will work equally well for the teleological approach. Though the subsequent discussion will presuppose the truth of the former account, it will not be parochial—the arguments could be recast in the teleological idiom for those who prefer it.

Note that one might advance a version of Bernard Williams' 'one thought too many' objection against either of these views. When I bear some cost to advance my son's interests, for example, I do not justify that cost by appealing to the value of my relationship—indeed, to do so would seem inappropriate. I should assume that cost simply because he's my son, without the additional appeal to our relationship's valuable properties. I think this objection is mistaken. When considering costs to myself, it would indeed be churlish to keep score, like a friend who, after an evening of shared rounds, tots up what you paid and what he paid, and demands the difference from you. But when weighing duties owed to my son against commitments I owe to others, in particular when offering arguments to justify treating my son preferentially, the appeal to the value of the relationship is not one thought too many, but rather a necessary part of the justification. Simply saying 'he's my son' is not enough to justify my conduct to others.

The next step is to identify the properties of our deepest relationships that ground strong, direct reasons for action. My focus in this paper is exclusively on the deep relationships that are most commonly thought to ground associative duties—those between lovers, parents and their children, friends, and comrades-in-arms. The qualities of these deep relationships are perhaps immediately perspicuous. Nonetheless, it is worth dwelling on them briefly.

Although our most intimate relationships can sometimes cause unbearable sadness, particularly when they end, they also can bring great pleasure—both the fun and hilarity of good times, and solace and relief in times of difficulty. Part of being in a valuable intimate relationship is that each makes the other feel better. Each inspires in the other positive mental states. Besides these affective properties, there are also other more complex components. Take, for example, the special understanding of one another possible in our deepest relationships. This understanding has at least two dimensions, one grounded in shared experience, the other in openness. For

example, my wife knows me better than anyone else does because she has shared my experiences, living through them with me. She knows every challenge that I have faced, every triumph, every defeat, every success or embarrassment; trivial and significant. She has watched me mature as we have become parents. She has inspired and influenced me, and in doing so has shaped my choices, as I have shaped hers. Our lives are intertwined. Nobody could know me as she does, because nobody has shared my life as she has.

As well as in shared experience, understanding derives from openness. In everyday life we all (except in Camus novels) adopt different personae in different contexts. It is only with those closest to us that we present ourselves whole, or at least nearly whole, without the veneers we habitually erect when among strangers. Openness and shared experience are constitutive parts of our most intimate relationships, and engender a special form of communication particular to those relationships and not present in our dealings with strangers or acquaintances. They are a bulwark against solipsism and existential angst. Intimate relationships are deep sources of value and meaning. They are worth protecting.

Besides giving us joy, solace, and the sense that we are not alone in the world, our deepest relationships may also inspire what is best in our characters. While there are many historical examples of extraordinary altruism wherein one person has made great sacrifices for strangers, these acts tend to be performed only by an exceptional few. Extensive self-sacrifice for one's loved ones, by contrast, is commonplace. This is more obviously true in poorer countries, where millions of men and women endure long days in unsafe factories for pitiful wages to provide for their families. For every family that has been lifted out of poverty, there are parents who have struggled their whole lives to give their children a better future. Even in the comparative comfort of modern liberal democracies, we still face difficulties; not least the universal problem of grief. We show strength for one another in the hardest of times, holding back our own fears and despair so that we can look after our loved ones. Many parents still spend a lifetime working for their children. Many sons and daughters do the same for their parents, in their old age. Our deep personal relationships are sites of the noblest and bravest deeds that most of us ever do. This too is worth protecting.

I will not try to define love, except to say two things. Love probably comprises the parts already discussed—the joy and solace, understanding and openness, devotion and sacrifice; it probably also amounts to more than these things. It is one of the worthiest things of which we are capable.

These properties of our deepest relationships are eminently deserving of protection—they give strong, direct reasons for action. Of course, people will often have relationships with friends, family, and lovers, which evince few of these properties. Close relationships can be the site of what is worst, as well as what is best, in human nature. My argument is confined to the good exemplars of our deep personal relationships.

The fellowship of arms could simply be folded into the broader discussion of deep personal relationships—many of the properties just described, in particular the mutual affection and shared understanding, are also fundamental to the relationship between comrades-in-arms. However, the significance of the fellowship of arms to the self-understanding of participants in combat (see the Introduction) makes it worth addressing on its own. Despite this, its role in the ethics of war is rarely discussed. There seem to be three reasons for this. First, while combatants often do form powerful bonds with one another, they also seem uniquely adept at bullying some of their number (e.g. Norton-Taylor, 2006). Second, and relatedly, there is undoubtedly a risk of over-romanticising this fellowship. One thinks of the solemn protestations of love in Erich Remarque's *All Quiet on the Western Front*, the eulogising in Wilfred Owen's 'Anthem for Doomed Youth' and the clichés that dominate Army recruitment advertisements (O'Brien, 1991, p. 79; Owen, 1996a, pp. 193-4; Remarque, 1996, p. 66). Finally, fellowship, or at least a desire to conform to the standards of the group, was a notorious factor in the commission of atrocities in twentieth century wars, in particular Vietnam (Bourke, 1999, ch. 6). We must remember, then, that the reasons that good relationships with one's fellow combatants can give derive from the specific properties of specific good relationships—not from all relationships between all comrades-in-arms, irrespective of these details.

What properties of special relationships between comrades are worth defending? We can identify at least three key features: shared experience, self-sacrifice and teamwork.

Combatants who have trained and fought together have shared some of the most challenging moments in their lives (Ninh, 1993, p. 217; Barnham, 1975, p. 26; Gray, 1998, p. 27; O'Brien, 1991, p. 194; Owen, 1996b, pp. 197-8; Remarque, 1996, p. 97). They can empathise with one another in a way that they cannot with people who have not had these experiences. It is important to recall the extent to which the conduct of war differs from ordinary life. Most of us have never even been in a fistfight, let alone a situation where a large number of people are concertedly trying to kill you, and you

them. The experience of war is life-altering. Combatants share these experiences, and can correspondingly achieve a unique mutual understanding.

Combatants who fight together also routinely perform acts of courage and self-sacrifice for the sake of one another unlikely to be replicated in ordinary life (Ninh, 1993, pp. 178-9; Gray, 1998, p. 46; Lee, 2006, p. 42; Nicol, 2007, pp. 114, 214; Remarque, 1996, p. 97). They take huge risks with their own lives to save others. This loose reciprocity of heroism is unique to the context of combat. The scale and pervasiveness of lethal threats make individual courage and self-sacrifice a necessary condition for group survival.

Lastly, combatants who fight together are part of a team, carefully organised to achieve common goals (Gray, 1998, pp. 41ff). For many people, being part of a team affords a sense of identity, purpose, and power: as a team, they can achieve far greater feats than they could ever do separately.

This, then, is an indicative, but by no means exhaustive, list of some properties of our deepest relationships, which I suggest can ground direct reasons for action: joy, solace, affection, mutual understanding, shared history, self-sacrifice and teamwork.⁴ The next step is to argue that seeing these relationships as grounds of some particular duties counts as responding appropriately to those properties. In other words, recognition of which duties is either constitutive of the relationship, or a necessary condition of it obtaining?

My proposal is this: suppose A and B share a valuable relationship of the kind described; for A to respond appropriately to the valuable properties of that relationship, he must give greater weight to B's interests in his deliberations than if their relationship did not obtain. That is precisely what it means to view a relationship (and one's associate) as special. The special relationship acts as a moral amplifier.

But what does it mean to say that duties are amplified? The strength of duties varies along at least two dimensions: the costs they can justify imposing on the duty-bearer; and their relative weight when they clash with other moral reasons. We can call these dimensions stringency and gravity respectively (Frances Kamm calls them the 'efforts' and 'precedence' standards: Kamm, 1996, pp. 313, 321). In my view, special relationships amplify duties along both dimensions: in virtue of his relationship with

4. The natural objection that immediately arises is: what if these properties are generated by a relationship that is predicated on wrongdoing? Nazi death squads shared formative experiences, worked together, probably even risked their own lives to save one another. Did they have associative duties to one another, with the same weight as associative duties owed by just combatants to their fellows? I address this objection in detail below.

B, A must bear greater costs to serve B's interests than he would if there were no relationship: his duties to B are more stringent. And in virtue of that relationship, A's duties to B are more likely to override other moral considerations than if they were not owed to an associate: his duties to B are graver.

The next section focuses on substantiating the second claim. Here I will illustrate the stringency claim. Suppose that A is at the beach and sees B struggling in the water, looking likely to drown. If A had no connection with B, then he might be required to take on x cost in order to save him. But if B is someone with whom A shares a valuable relationship—his son, say—then the cost that he ought to bear will be greater than x . So, suppose B is caught in a rip, and A judges that he would be risking his life to try to save B. This would not be morally required were B a stranger (suppose), but may be required if A is his father, and they have a valuable relationship. This will of course depend on further details—A's prospects of saving B must be sufficiently high, for example—but the basic point should be clear.

Sometimes this amplification is best described by saying that A's duty to B is more stringent if they share a valuable relationship; sometimes the amplification results in a duty obtaining which would otherwise be absent. Special relationships might generate other duties which cannot be analysed in this way, but these amplifying duties are sufficient for the purposes of this paper.

What explains this amplification? Suppose you fail to perform a duty owed to your associate which you would have owed even in the absence of the relationship. You retain the reasons which apply in virtue of your shared humanity—you have damaged the interest protected by the duty, and disregarded the victim's moral standing (see Lazar, 2009b). But you have also betrayed a friend and disregarded the value of the relationship between you. This additional reason amplifies the force of the reasons you already have.

Acknowledging these duties constitutes an appropriate response to the valuable properties of our deep personal relationships. In the terms of the teleological account, these duties can be construed as constitutive of the relationship, since if you do not give your associates' interests greater weight than those of strangers, then how can you call this a special relationship? They might also be necessary conditions of the relationship obtaining—not least when our associates need to be protected against potentially lethal threats.

The next task is to show that the threat of war can engage specific duties that we owe to our family, friends, and comrades-in-arms to protect them from harm. Later in

the paper, we will consider precisely how these duties are transferred to combatants fighting on our behalf; here I show that these relationships can ground moral reasons that bear on the justification of fighting.

Often our duties to protect our associates will weigh *against* fighting: we best protect our associates by trying to stop the war altogether (even, sometimes, if it would otherwise be justified). This is an important and plausible restriction on the Associativist Account. But it should not be overstated. Sometimes the onset of war is inexorable, and cannot be halted by comparatively uncoordinated individuals—perhaps because the adversary is implacable, or because our own leaders are hell-bent on conflict. When we can do nothing to prevent war, our duties to protect our associates may enjoin fighting.

This is most obviously true for active-duty combatants, whose vital interests are guaranteed to be at stake once fighting begins. When combatants are in harm's way, their special relationships can at least sometimes ground strong, direct moral reasons to protect one another from those pervasive threats. Recognising a duty to protect your comrades-in-arms from lethal threats is a necessary condition of those relationships persisting. It is also arguably constitutive of fellowship of arms that you take risks to ensure no-one gets left behind.

War's effects are often felt more severely by civilians than by combatants. This is easy to forget, at least for citizens of the UK, US, and our allies, since we fight our wars far from home. For our adversaries in those wars, however, the story is often quite different. For other liberal democracies such as Israel and India, war directly threatens the lives and vital interests of ordinary civilians. In conflicts in and between developing nations, it is axiomatic that the greatest suffering falls on those least able to defend themselves. Indeed, it is often argued that 90% of the victims of warfare are noncombatants; Adam Roberts has recently disputed that figure, but his objection is more to the project of identifying civilian/military casualty ratios, than to the thesis that civilians suffer inordinately from the effects of war (Roberts, 2010).

Hugo Slim offers this depressing catalogue of civilian injuries endemic to armed conflict: 1. Killing, torture, wounding; 2. Rape and other forms of sexual violence; 3. Forced and restricted movement; 4. Impoverishment; 5. Famine and disease; 6. Emotional suffering caused by loss; 7. Post-war suffering due to UXO as well as psychological problems of post-traumatic stress (Slim, 2007, p. 39). It should be uncontroversial that, if we can protect our loved ones against harms such as these, even at considerable cost to ourselves, we ought to do so. Sometimes the best way to avoid

this suffering is to oppose the war before it begins. Once that option is gone, however, fighting may be the only available means to protect those we love.

Besides these threats to individuals' vital interests, war also threatens other social goods such as our state's territorial integrity, the character of our political institutions, and the stable cultural preconditions of a good life. Sometimes the threats to vital interests arise only because we use force to avert these threats to social goods—these are cases of purely political aggression (Lazar, Forthcoming-a). Do we have associative duties to avert threats to these social goods? Perhaps if we owe associative duties to our compatriots, *qua* compatriots, we might do so. But in this paper my focus is on our deeper personal relationships, and it is less plausible that we owe it to our children, for example, to protect our democratic institutions. Such a duty is much less obviously constitutive of, or necessary for, our deep personal relationships. As such, the ensuing discussion will focus exclusively on threats to our associates' vital interests. This means that the Associativist Account, as presented here, cannot justify going to war against an adversary that threatens only these political and social goods.

THE GRAVITY OF ASSOCIATIVE DUTIES

I have argued that we respond appropriately to the valuable properties of our deep personal relationships by giving our associates' interests greater weight in our deliberations than would be justified absent that relationship. This means bearing greater costs to protect their interests than we are required to bear for non-associates. In this section, I defend the more controversial view that we can justifiably impose greater costs on non-associates to protect our associates' interests than would be permissible without that relationship. Indeed, I argue that we can permissibly kill non-associates, to save the lives of associates.

To begin with, however, it is helpful to consider the infliction of harms short of death. Suppose a meteor is plummeting from the sky, directly towards B, and will surely kill her if nothing is done. A can deflect the meteor using her surface-to-air heat-seeking missile launcher. But if she does so, fragments of the meteor will hit a nonliable person C. Suppose, for now, that A bears no relation to either B or C. She must consider at least two dimensions of her actions. If she lets the meteor fall towards B, then she must consider the harm that B suffers and the qualitative evaluation of her agency in letting B suffer that harm. If she diverts it towards C, she must

consider the harm that C will suffer and the qualitative evaluation of her agency in diverting that harm towards C. Her two options, in other words, each involve qualitative and harm-based dimensions. Suppose that intervening in a causal process leading to harm is qualitatively worse than failing to prevent a causal process from resulting in harm. If the prospective harms to B and C are equal, then A ought to do nothing, since letting B die is better in one respect and in no respect worse than killing C. However, if the prospective harm to C is less than the harm to B, then A might be required to divert the meteor—though the qualitative difference between her options means that the harm to C must be more than marginally less than that to B. But there is some threshold of harm H, such that if the harm to C is less than H, A is morally required to divert the meteor.

I suggest that when B and A share a valuable relationship, the threshold of harm A can permissibly inflict on C to save B's life is higher than if A and B are not associates. This is so for two reasons. First, in virtue of their valuable relationship, A must give greater weight to B's interests than would otherwise be so. Second, the agential dimension of failing to save B is qualitatively worse given their relationship—it amounts to a failure to recognise the value of their relationship. A has weightier reasons to save B's life when they share a valuable relationship, so those reasons can outweigh more harm inflicted on C. In the terms introduced above, A's duty to protect B is graver when B is her associate.

This view contravenes philosophical orthodoxy, according to which our associative duties cannot override general negative duties (for further arguments against that orthodoxy see Lazar, 2009c). One way to resist that orthodoxy is to dispute the significance it ascribes to the positive/negative duty distinction. I do not pursue this option; but we should note that this distinction is one of degree, not of kind. Breaching negative duties is not lexically worse than breaching positive duties—as though we ought to perform any negative duty, no matter how slender the interests at stake, in preference to any positive duty, no matter how serious. And as the meteor example shows, sometimes it is permissible to breach negative duties as a lesser evil, even in the absence of a special relationship. My contention is simply that our deep personal relationships affect how much harm can be inflicted as a lesser evil.

One might concede this much, yet deny that our associative duties can justify *killing*, even if they could justify breaching lesser negative duties. However, while killing a nonliable person is clearly a presumptively wrongful act, not all acts of killing are alike. There are important agential differences between killing and letting

die, between intentional killing, unintended but foreseen killing, unintended, unforeseen, but foreseeable killing, and unintended, unforeseen, unforeseeable killing. There is a further distinction between eliminative and opportunistic killing. In eliminative killing, the killer derives no benefit from the victim's death that he would not have enjoyed in the victim's absence. In opportunistic killing, the killer does derive such benefits (Frowe, 2008; Quinn, 1989; Quong, 2009; Tadros, 2012). If A diverts the meteor towards C, then while A clearly foresees C's death she does not intend it. Her killing is also eliminative, insofar as she would be no worse off if C were elsewhere. Contrast this with an alternative, where unless the meteor lands on someone, it will explode and kill B anyway. If A diverted the meteor towards C in this case, she would be using C opportunistically, deriving a benefit that would have been unavailable in her absence. This would be a qualitatively worse killing.

At least in 1:1 cases, the duty to protect one's associates from lethal harm cannot override the general negative duty not to intentionally, opportunistically kill a nonliable person. But, as I now argue, it can override the general negative duty not to foreseeably eliminatively do so.

My argument that A's associative duty to protect B can justify killing C starts with the comparison of three cases, and a commitment to transitivity in ethical reasoning, according to which if the moral reasons for ϕ -ing outweigh those for ψ -ing, and those for ψ -ing outweigh those for μ -ing, then the moral reasons for ϕ -ing outweigh those for μ -ing. The transitivity of moral reasons has in recent years come under sustained criticism from Larry Temkin and Jonathan Dancy, among others (Dancy, 1993, 2004; Temkin, 2012). Though I cannot address their arguments here, transitivity remains the default, commonsense position. To deny transitivity in some case, you need an account of precisely why it fails in that particular instance. I consider one such account below.

Case 1: A meteor is plummeting towards the earth, and, if A does nothing, will kill five nonliable people. If she uses her missile launcher, however, she can divert it away from the five. If she does so the meteor will land on and kill C.

This is just the standard trolley case, albeit with meteors instead of trams. I am assuming, note, that the cost to each person of dying is the same, and that the costs to others of their deaths are also the same. It is intuitively clear to most that A is at least permitted to divert the meteor. Indeed, anyone who denies that it can be permis-

sible to foreseeably kill an innocent person as a side-effect of saving others (or who thought the ratio had to be much higher than 5:1) would have to be a pacifist. Wars cannot be fought without foreseeably killing innocents, so this quasi-absolutist attitude would make fighting a justified war impossible. I think the right interpretation of case 1 is that A *ought* to divert the meteor. The combination of the five's interests, and the agential component of letting them die, outweigh C's interests, and the agential component of foreseeably, eliminatively killing her.

Case 2: Two meteors are plummeting towards the earth. One is headed towards a group of five, as in case 1. The other is headed towards A's daughter, B, with whom A shares a paradigmatic special relationship. A has only one missile, and can divert only one meteor. The diverted meteor will land harmlessly in a field.

Again, I think it is intuitively obvious that A is permitted to save B rather than save the five. Moreover I think that A is morally required to save her daughter, rather than save the five—she would be acting wrongly if she did not. I offer more discussion of this distinction below.

Case 3: There is only one meteor again, and it is headed towards A's daughter B. A can divert the missile, but it will subsequently land on and kill C.

I think A ought to divert the meteor, killing C, in order to save her daughter. The argument to reach this conclusion can be stated in two versions, one stronger than the other. While I think the stronger one is right, it depends on more contentious intuitive judgments than the weaker version, so I will concentrate on defending the latter.

The strongest form of the argument is this: in case 1, A is morally required to save the five, rather than avoid killing C; in case 2, she is morally required to save her daughter, rather than save the five; so, by transitivity, in case 3 she is morally required to save her daughter, rather than avoid killing C.

The weaker version is this: in case 1, A is permitted to save the five, rather than avoid killing C; in case 2, she is permitted to save her daughter, rather than save the five; so, by transitivity, in case 3 she is permitted to save her daughter, rather than avoid killing C.

Why should we believe the stronger version of this argument? Some of course

will simply share my intuition about case 3, so the argument from transitivity will be otiose. Those who do not initially share that intuition, but do agree with my assessments of cases 1 and 2, might be persuaded to change their minds about 3 if they think that A's deliberations in these cases should take the form of weighing her reasons for action to reach a conclusion about what she ought to do, all things considered. We can model this by using numerical values to capture the weight of her reasons—remembering, of course, that the numbers are just a heuristic. Following the model introduced above, we can distinguish between agential reasons against ϕ -ing, and harm-based or, better, interest-based reasons against ϕ -ing. We could either posit that agential reasons are multipliers, or other more complicated functions, or that agential and interest-based reasons are additive. For simplicity, let us use the latter approach.

Suppose the interest-based reasons against killing C (independent of the agential dimension) are of magnitude 100, while the agential reasons against killing her are of magnitude 900. The interest-based reasons against letting each of the five die are also of magnitude 100, while the agential reasons are, for each one, 110 say. A's reasons to save the five amount to 1050, while her reasons not to kill C amount to 1000, so she ought to save the five by killing C.

On this model, if A is required to save B rather than save the five, then the magnitude of her reasons for saving B should be greater than her reasons for saving the five. On my view, A ought to give her associate's interests additional weight in her deliberations. Plus, there is a distinct agential dimension to failing to protect someone she shares a valuable relationship with. We could model this by saying that her interest-based reasons to save B amount to 200, while the agential reasons against letting her daughter die amount to 900, giving a total of 1100.⁵

The specific numbers, of course, are arbitrary; they are just a means of modelling A's deliberations. But if this is a sound way for A to deliberate, then the permissibility of her killing C to save B should follow—if her reasons to save B are greater in magnitude than her reasons to save the five, and her reasons to save the five greater in magnitude than her reasons not to kill C, then, since 'greater in magnitude than' is a transitive relation, her reasons to save B should be greater in magnitude than her reasons not to kill C.

5. Suppose each of the five also has valuable relationships with their children, which will obviously be terminated when they die. Should A take these into account? I think not—the reasons relationships give are agent-relative, and can only be transferred to another when that other is both authorised by the relevant agent, and accepts that authorisation, as I discuss in the next section.

However, this model supports the stronger version of the argument only if A is indeed morally required to save B rather than the five, in case 2. And while I think A is definitely not required to save the five rather than B, I do not have a conclusive argument that she is required to save B rather than the five.

Morality cannot plausibly be so demanding as to require A to sacrifice her chance to save her daughter, in order to save the five. The central theme of much nonconsequentialist ethics—apparent, for example, in the contrast between eliminative and opportunistic agency mentioned above—is that there is something inherently objectionable about our being made into resources for the advancement of valuable states of affairs, or opportunities for others to exploit to their advantage (for a powerful recent statement of this classic view, see Tadros, 2012). Just as A may not use C as a means to save five, she cannot be required to make herself into a means to save them, when the cost of doing so is as great as sacrificing her daughter's life.

I am less certain that A is morally required to save B, however. Though I think she certainly wrongs B by failing to save her, she may not act wrongly, all things considered. On my view of associative duties, B has a justified claim against A, grounded in their valuable relationship, that A give her interests greater weight than if their relationship did not obtain. By failing to save her, A fails to give B's interests that weight and shows disrespect for their relationship, wronging her. B has justified grounds for complaint against A. To deny this, while accepting the argument of the previous paragraph, would be to suggest that the moral significance of A's duties to B is exhausted by their contribution to A's well-being. This would effectively reduce those duties to prudential reasons. This is a serious category mistake: our reasons to protect our nearest and dearest are not merely complicated reasons of self-interest. Failing to protect those you care about is not merely irrational or stupid—it amounts to a genuine moral failing.

However, while A wrongs B by failing to save her, there are two interpretations consistent with this judgment. On the first, A simply acts wrongly, all things considered. The second is that, in saving the five, A permissibly breaches her duty to protect B in order to do something supererogatory (Kamm, 1996, p. 317). Though saving the five is all things considered permissible, it involves *pro tanto* wronging B. If A saves B, then she both fulfils her associative duty to B, and acts merely permissibly. This interpretation captures my view that A genuinely owes B a duty to protect her—it is not merely a question of fulfilling her self-interest—while remaining consistent with the conclusion that A is merely permitted to save B, rather than required. And as Frances

Kamm has argued, sometimes supererogatory acts can render it permissible to breach weighty duties—she gives an example where you break an appointment in order to give one of your kidneys to the victim of a traffic accident.

My theory and intuitions strongly favour the first interpretation: A acts impermissibly if she fails to save her daughter. By saving the five, and abandoning her daughter, she is not showing an admirable capacity for self-sacrifice, but an objectionable lack of regard for her child, and the relationship they share. In my view she would look like the Dickens character, Mrs. Jellyby, whose concern for the distant needy leads her to neglect her own children. Suppose the two meteors are each heading towards a flash drive, and on one is the only copy of my book manuscript; on the other is the only copy of your book manuscript. If I decide to divert the second meteor, saving your book, then it seems that I have done something generous, supererogatory and praiseworthy. Work on my book was a crucial part of my well-being, and I have sacrificed that for your sake. Matters are quite different when it is the chance of saving my son that I have sacrificed, in order to save somebody else's (thanks to Simon Keller for this example).

However, one could perhaps interpret the case the other way, and argue that while it would perhaps be wrong to save one stranger rather than her child, when there are five other lives on the line, it is permissible for A to sacrifice her chance to save her daughter, for the greater good. Those who agree with the first interpretation of the case should be satisfied with the stronger version of the argument; to convince those who affirm the second interpretation, I need to defend the argument in its weaker form.

The weaker version of the argument, however, is vulnerable to the objection that permissions do not obey transitivity. We can adapt a trio of examples from Frances Kamm to bring out the relevant worry:

Case 4: A has an appointment to meet C at 12 o'clock. On the way there, however, she sees a child drowning in a river. Suppose that the risk to A of saving the child is sufficiently great that doing so is supererogatory—she is not morally required to save the child. Nonetheless she decides to do so, and as a result misses her appointment.

It seems intuitively obvious that A has acted permissibly. But now consider

Case 5: A is walking past the same river on another day, on the way to hit some golf

balls at the driving range. Again, there is a drowning child, whom A could save only at the same risk as in the previous case. A heads to the range.

Here it is clearly permissible for A to continue on to the driving range, rather than save the child. After all, saving the child is by hypothesis supererogatory, so A cannot be required to do it. But now consider

Case 6: A is heading for her appointment with C, but decides to go and hit some balls at the driving range.

Clearly A acts impermissibly in this case. And this despite the fact that she was permitted to go to the driving range rather than save the child, and permitted to save the child rather than keep her appointment. Nonetheless she is not permitted to go to the driving range rather than keep her appointment. So the transitivity of permissions appears to fail.

One might think my cases 1-3 are analogous, respectively, to cases 4-6. In case 1, A is permitted to save the five, at the cost of breaching her obligation not to kill C; similarly, in case 4 A is permitted to save the drowning child, at the cost of breaching her obligation to meet C. In case 2, A is permitted to save B at the cost of not saving the five, because saving the five would be supererogatory. In case 5, A is permitted to play golf at the cost of not saving the child, also because saving the child would be supererogatory. So, just as transitivity fails for case 6, where clearly A is not permitted to play golf at the cost of breaching her obligation to keep her appointment, it should also fail for case 3, and A should not be permitted to save B at the cost of breaching her obligation not to kill C.

Stated this baldly, the objection fails. The appearance of intransitivity disappears once we properly describe the reasons applying to A in cases 4-6. A is permitted to break her appointment in case 4 because saving the drowning child matters more than keeping her appointment. The costs to her of doing so are deliberately irrelevant—though sufficient to make saving the child supererogatory, this does not impact on the permissibility of her action (other than to entail that she could permissibly keep her appointment). In case 5, the costs to A of saving the child are deliberately central—but her interest in playing golf is irrelevant to what she is permitted to do. Saving the child is supererogatory because of the sacrifices involved in attempting the rescue—the risk to A's life, etc. Given those risks, A is not required to save the

child and can use her time as she wants—so why not play golf? But the permissibility of her trip to the driving range is independent of her interest in getting there. Finally, in case 6 A weighs her obligation to C against her interest in playing golf. The objection is that, if transitivity held, we should be able to infer her conclusion from cases 4 and 5. However, these examples tell us nothing about transitivity, because A's deliberations in cases 4 and 5 are different, and different again in case 6. Why should A's interest in playing golf override her appointment to meet C, just because she is permitted to save the child rather than keep her appointment, and permitted to play golf rather than save the child, when the reason she is permitted to play golf rather than save the child has nothing to do with her reasons for playing golf, and everything to do with the risks of her own life of rescuing the child?

Case 5 licenses conclusions about case 6 only if the costs invoked are the same in each case. Suppose that the costs to A in case 5 of saving the drowning child were *just* that she miss out on her trip to the driving range. If that were so, clearly she would be required to save the child, just as she is required to make her appointment, even if doing so means missing out on practising with her 9-iron. Conversely, suppose that the costs in case 6 were the same as those in case 5: to make her appointment, A has to undergo just the same risks as are involved in saving the drowning child. Keeping the appointment would then be sufficiently costly that doing so is supererogatory and A would be permitted to do what she likes with her time.

This discussion should indicate how cases 4-6 are crucially disanalogous from cases 1-3, in which the costs *are* consistently invoked across cases. Adopting the interpretation given above, according to which A is merely permitted (not required) to save her daughter rather than the five, the key point is that in cases 2 and 3 the costs to A are identical: foregoing the opportunity to save her daughter. If those costs are serious enough that not saving the five is permissible, and if saving the five matters more than not killing C (as per case 1), then shouldn't those same costs also matter more than not killing C?

However, while this deals with first objection to the weaker argument, it tees up a much more serious second objection, which argues that my purported transitivity fails because the first two cases do not give enough information to draw conclusions about the permissibility of killing C in case 3. Case 1 tells us that the duty to save five is graver than the duty not to kill C, and case 2 tells us something about the stringency of the duty to save five, but nothing about the gravity of the associative duty to

save B. So, how can we draw warranted conclusions about either the stringency of the duty not to kill C, or the relative gravity of that duty and the duty to save B?

One might respond that the duty not to kill C is less stringent than the duty to save the five just in case the latter duty is graver than the former. That would be enough to save my argument; unfortunately it is quite implausible, as the following trio of cases shows.

Case 7: Two meteors are falling towards the earth, each headed for a group of five people. A can press a button that will fire missiles at each meteor, diverting them. But this will lead to the meteors instead crushing two different people ($C_{1,2}$).

Case 8: There are three meteors plummeting towards the earth. Two are as in case 7, each heading for a group of five people. One is heading for A. She can press a button that will fire a missile that will divert the meteor headed at her, or one that will fire two missiles, saving the ten. Either way the diverted meteor(s) will land harmlessly.

Case 9: There is only one meteor again, and it is headed towards A. A can divert the missile, but if she does so it will land on, and kill, two people ($C_{1,2}$).

I assume that if diverting the meteor is permissible in case 1, it is permissible in case 7. This could involve the fallacy of composition, but probably does not: if the duty to save five lives is graver than the duty not to kill one, then two instances of the former duty should be graver than two instances of the latter. I also assume that A is permitted to save herself in case 8, even though it means letting the ten die. So the duty to save ten is graver than the duty not to kill two, but not stringent enough to require A to sacrifice her life. If duties' relative stringency can be inferred from their relative gravity, then in case 9 A should be permitted to divert the missile towards $C_{1,2}$. Her duty not to kill them cannot be more stringent than her duty to save the ten, since it is less grave, as case 7 shows. But it is surely not plausible that A is permitted to kill two to save herself.

To save the weaker version of the transitivity argument, I need to do one of two things: either show how facts about stringency can be inferred from facts about gravity; or argue that the permission to save B rather than the five is more than a reflection on the stringency of the duty to save the five. Recall that, since the stronger version relies only on gravity to make its case, it is untouched by this objection.

I think it would be unusual for stringency and gravity to be either wholly disconnected from one another, or connected by some sort of discontinuous relation. The considerations that ground a duty's weight relative to other moral reasons, and those that ground its weight relative to the interests of the duty-bearer seem at first sight to be just the same considerations. When we assess the gravity of the duty to save another person, we adduce the moral weight of that individual's interests, and the independent significance of her standing relative to the duty-bearer (*qua* human being, and/or *qua* associate); when we assess that duty's stringency, we appeal to just the same considerations.

However, one consideration does bear on stringency and gravity in different ways. This is the basic nonconsequentialist thesis adduced above; that there is something inherently objectionable about our being made into resources for realising valuable states of affairs. Call this the means principle. The means principle grounds important differences in the stringency of positive and negative duties. If A were morally required to sacrifice her own life to save the ten in case 8, morality would turn her into a means for the realisation of the best state of affairs, at the cost of her own life. She makes herself into a means in two senses: first, she is literally the means whereby the threat to the ten is averted (she fires the missiles that divert the meteors); second, she thereby provides the ten with a benefit that they could not have enjoyed in her absence. If A chooses to sacrifice herself, then there is no special problem here: we are clearly permitted to make ourselves into means to the greater good if we choose. But if A would rather save herself, for morality to tell her to do otherwise is to make her into a means, which is inherently objectionable. Of course, sometimes the good we can achieve by being made into a means is sufficiently great that it overrides this inherent objectionability. But the means principle diminishes the stringency of A's positive duties independently of the considerations adduced above. Where the duty makes A into a means to bring about the good, the costs it can require A to bear to achieve that good are reduced by a function M, which captures the distinctive objectionability of her being made into a means.

Negative duties, by contrast, do not make us into a means for realising the best states of affairs. If A performs her duty not to harm C, in case 3, she does not become a means. She does not thereby save C. Nor does her presence provide C with a benefit she could not have enjoyed in A's absence. Had A not been there, C would have been just fine. Since adhering to negative duties does not make us into a means, there is

nothing inherently objectionable about them. Accordingly, the costs that we must bear to perform those duties are not reduced by the M function.

The means principle explains, then, why a given positive and negative duty can be differentially exacting, even when other considerations that ground stringency and gravity are basically the same. And importantly, while the means principle has some bearing on the gravity of positive and negative duties, it does so in quite a different way: it lends additional gravity to duties not to turn others into a means. But while such duties are more likely to be negative than positive, they can also be positive, and plenty of negative duties are not grounded in the means principle. One salient example is the negative duty not to harm a nonliable person eliminatively, with which we are primarily concerned here.

So, even if the other reasons that ground our assessment of a positive and negative duty's relative gravity are identical, they can be differentially stringent because of the M function's contribution to the stringency of positive duties. Thus we can consistently affirm (in cases 1-3) that it is permissible to save the five rather than not kill the one, permissible to save B rather than save the five, and yet impermissible to save B rather than not kill the one.

We could perhaps learn something about duties' relative stringency from their relative gravity. Let A be the agential reasons grounding a positive (PD) or a negative duty (ND); let I be the interest-based reasons; and let M be the means principle function. When establishing the relative gravity of a PD and a ND, where > means 'is graver than'

$$PD > ND \text{ iff } (A_{PD} + I_{PD}) > (A_{ND} + I_{ND})$$

Whereas, when establishing the relative stringency of a PD and a ND, where \gg means 'is more stringent than'

$$PD \gg ND \text{ iff } (A_{PD} + I_{PD})M \gg (A_{ND} + I_{ND})$$

So if we knew the value of M, and the amount by which $(A_{PD} + I_{PD})$ exceed $(A_{ND} + I_{ND})$ in the gravity calculation, we could infer that $PD \gg ND$. This could, in theory, vindicate something like the weaker transitivity argument. However, I doubt we can confidently define the M function; indeed I suspect that it is not simple and continuous. We could perhaps ask when we are required to sacrifice our lives to save

others', but I lack an intuitive grasp on that question. On the one hand, it seems plausible that I shouldn't be required to sacrifice myself unless by doing so I can save many thousands of others' lives. On the other hand, this could just be selfishness, and morality might really be more demanding than that, so if I was certain that I could save thirty, or maybe even twenty equally good lives, I ought to sacrifice myself. Unable to confidently choose between these intuitions, I think we should abandon the attempt to set a value on *M*, and concede that this avenue to saving the weaker argument from transitivity is at best incomplete.

The second approach, recall, is to argue that the permission to save *B* rather than the five is more than a reflection on the stringency of the duty to save the five, hence cases 7-9 are not in fact analogous to cases 1-3. This is because *A*'s permission to save *B* depends crucially on the duty she owes to protect *B* from harm, not merely on the great cost to her of not saving *B*. Since cases 8 and 9 involve only costs to *A*'s well-being, rather than any associative duties, we can consistently deny that killing C_{1-2} is permissible in 9, while affirming that killing *C* is permissible in 3.

My own view, of course, is that even this interpretation of *A*'s position is not quite right—I think that her duty to protect *B* is graver than her duty to protect the five, which is graver than her duty not to kill *C*, so her duty to protect *B* is graver than her duty not to kill *C*. In other words, I affirm the strong interpretation of the argument. But surely some who reject that view will find this intermediate alternative attractive: there is something fundamentally odd about supposing that the deliberative force of *A*'s reasons to protect *B* are reducible, in case 2, to the contribution *B* makes to *A*'s well-being. This paper is premised on the assumption that our associative duties are genuine moral reasons. If that is true, then it is possible that only when the costs to *A* are combined with her duties to *B* is it permissible for her not to save the five, and likewise in case 3—the moral reasons grounded in *A*'s associative duties to *B* are a necessary condition of her being permitted to kill *C*.

The argument that *A* can permissibly kill *C*, in the course of saving *B*'s life, should go through for those who agree with the strong interpretation, according to which *A* is required to save *B*, at the cost of *C*'s life; and for those who favour the intermediate interpretation, which says that *A*'s associative duties to protect *B* are a necessary component of the permissibility of her killing *C*. Those who affirm the weak interpretation may remain unconvinced. For them, one last gambit is possible. At least some will believe that it is permissible for *A* to kill *C* (in the manner described in the cases above), when she has no other means of saving her own life. That

is, they think the duty not to kill C is not stringent enough, in such cases, to require A to sacrifice her own life rather than take C's. These philosophers should agree that, on the same grounds, it is permissible for A to kill C in the course of saving B in case 3, at least if A would choose to sacrifice her own life to save B, if she could. For if A would make that sacrifice, then she regards death (while B lives) as preferable to life without B. If she cannot be required to bear the cost of death to avoid killing C, then she cannot be required to bear the comparable cost of life without B. This argument from relative costs should persuade at least some who reject both the strong and the intermediate interpretations of these cases.

OPERATIONALISING ASSOCIATIVE DUTIES

Suppose that war is imminent and that we can do nothing, as individuals, to prevent it: either because we face an implacable adversary, or because we cannot alter our government's course of action. We and our closest associates are under threat; necessarily for active duty combatants, most likely for noncombatants too. Our loved ones face the prospect of grievous damage to their most vital interests. Each of us owes it to our associates to protect them, if we can, against these threats. I have just argued that these duties can override some duties not to harm others, but how are they operationalised, so that they become directly relevant to the justification of killing in war?

Combatants' duties to protect their comrades-in-arms are operationalised automatically by the onset of conflict—in defending their comrades and their associates back home, they are acting (at least) on reasons that apply to them directly. In a *levée en masse*, where erstwhile civilians rise up to defend themselves and their families, the same is true—each person can permissibly fight (and kill) while appealing only to reasons that directly apply to him. But this is not the whole story. I will argue that, in situations like those described in the previous paragraph, combatants can act on behalf of the community of which they are a part, and that their actions can be justified by reasons that apply directly to the members of that community. They act not only on reasons that apply to them *qua* individuals, but on reasons that apply to them *qua* representatives of that community, which include the associative duties that are, in part, the very grounds for living together in communities and organising ourselves for collective defence.

Imagine a village in the state of nature, in which 100 families live. Suppose that

the village lacks any institutional structure besides a few informal public spaces. One day, the village is attacked by a marauding band. The villagers, disorganised as they are, each defend their own families and friends from this murderous crew. It should be uncontentious, I think, that part of what justifies each person in using force to defend his loved ones are the associative duties that they owe to protect them. Perhaps one might think that, if all the marauders are liable to be killed, then the villagers have no need to appeal to their duties to protect their loved ones; simple principles of self- and other-defence will suffice. Suppose, then, that each marauder is on horseback, and each has an innocent human shield tied to the horse, so that the villagers cannot defend themselves without risking harm to innocents. Then the mere appeal to self-defence or other-defence is insufficient; we must invoke some duties which can override the duties not to harm innocents.

Suppose that they successfully fight off the band, but do not wipe them out. After they have treated the wounded, and cleared up the damage, a village meeting is called and the villagers decide that some among them—the most martially adept—should stand guard over the others. Some time after they establish this militia, the marauders come back. Surely whatever reasons justified defence in the first case, where the villagers each fought to defend their own, should also justify defence in the second case, where they have deputed some to fight on their behalf? Why should those reasons disappear? On one view, the whole justification for forming political communities—and for forcing outliers to join in—is to enable common defence. If the reasons in favour of common defence appeal to associative duties then why should those reasons disappear merely by virtue of the fact that the villagers have a more robust way to ensure that their loved ones are protected?

Warfare requires natural attributes of courage and skill, and considerable training. Many people would make ineffective soldiers. Moreover, these differences in ability raise serious problems of fairness. Those who can fight effectively are better able than others to perform their associative duties. Given that, as argued above, performing those duties is a fundamentally important part of our lives, this is cause for concern (for a similar argument, see Brighouse and Swift, 2009). Of equal concern, in most conflicts it will not be necessary for every able-bodied person to fight. Indeed, it would most probably be counterproductive. Thus non-fighters free-ride on the others' military capabilities, which in turn threatens serious collective action problems. We should resolve these problems by creating institutions to enable us all to perform our duties to protect—whether associative or general—in a fair and optimal

way. The armed forces are the executors of those duties: when soldiers fight, they are not simply responding to reasons that apply to them as individuals (though they are also responding to those). They are responding to reasons that apply to all of us in the community they represent, on our behalf. And when we assent to and provide for these institutions, we authorise them to defend our associates on our behalf.

This institutionalist argument operationalises the duties underpinning the Associativist Account, rendering them directly applicable in war and showing how the associative duties of civilians in a society can justify the killing done by their armed forces. It also forestalls some worries that the account might raise. For example, it allows us to remain agnostic on whether that duties to compatriots *qua* compatriots play an important role in justifying killing in war. Co-citizenship is quite different from the relationships that paradigmatically generate associative duties. Hence, many people are sceptical about duties to co-citizens. That scepticism may be misplaced, but the Associativist Account does not depend on proving that point. The armed forces exist for members of a political community to perform their associative and general duties to protect in a way that is optimal and fair. In most political communities, most people will share most of their special relationships with other people who are resident in the same territory. There are obviously exceptions, but the requirement of optimality justifies concentrating on those duties that substantially overlap.

This institutionalist argument addresses another worry about the Associativist Account. Participating in even a justified war can often damage combatants' valuable relationships. Kept apart from their families, and subjected to radically different experiences, rifts can emerge. If their participation in war were justified by reasons grounded in their own relationships, those reasons might in fact mandate not fighting, given the risk to their deep relationships.

This objection invites three responses. First, sometimes combatants ought to put their own special relationships first, and refuse to fight. Second, that injunction must be qualified by the countervailing pull of their duties to their comrades-in-arms. Given the relative threats faced, in some cases protecting their comrades-in-arms will be justifiable even if it means sacrificing their relationships with those back home. Third, combatants are not performing their own associative and general duties alone; they are also the executors of those duties for their whole society. These duties give them powerful reasons to serve, even when they undermine their own relationships by doing so.

The institutionalist move might draw some criticism from those who think that

our associative duties, as agent-relative reasons, cannot be transferred like this. My associative duty to protect my wife, for example, does not give you reason to protect her. That I could justifiably ϕ , because I have agent-relative reason to ϕ , does not mean that ϕ -ing is justified for you.

However, though agent-relative reasons do not automatically apply to other agents, under some conditions they can. If you are acting on my behalf—if I have authorised you to be my agent, and you have assented to that authorisation—then you should enjoy the same permission that would otherwise apply to my action. In virtue of my being permitted to ϕ , I have the moral power to authorise you to ϕ on my behalf, even though without that authorisation you would not be permitted to ϕ (see also Fabre, 2009, for a similar argument). The same dynamic operates in contracts. Suppose I have a right to collect a debt from you and I empower a bailiff to collect that debt. As he is acting as my agent, with my authorisation, he is entitled to respond to the reasons that apply to me. Without that authorisation, he would just be some guy trying to take your car.

I think that this objection confuses what Raz calls action reasons with agent-relative reasons (Raz, 1986, ch. 6). Action reasons are satisfied only if a specific agent acts on them. For example, I have a reason not only to ensure that my son is looked after, but to be the one who looks after him. When their lives are at stake, however, we surely have reason to ensure that our associates are protected, though not necessarily that we specifically be their protectors. The advantages of coordination are so great that it would be absurdly counterproductive for each to insist on carrying these duties out ourselves.

This shows how the duties that I owe to protect my son (for example) can be relevant to the justification of actions taken, on my behalf, by soldiers defending us against military aggression. The arguments above showed that—at least if one accepts either its strong version, the intermediate version, or the argument from relative costs—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. One might nonetheless think that, even if all this were true, the contribution made by the Associativist Account to justifying killing in war will necessarily be incomplete. After all, successfully fighting wars surely requires opportunistic and intentional killing, as well as eliminative and foreseen.

I readily concede that our duties to protect our associates cannot override duties

not to harm others opportunistically when the harms in question are otherwise roughly equal. The Associativist Account is not intended to be a comprehensive theory of the morality of killing in war, but to contribute one significant piece to that theory. I noted above that our reasons to protect our associates cannot justify defensive war against purely political aggression (which does not threaten our special relationships); I now also concede that they cannot justify the opportunistic killing that war inevitably involves. The best theory of the morality of war will need to appeal to more than just associative duties.

That the Associativist Account is incomplete does not, however, entail its irrelevance. Necessary killing in war covers the gamut, and some of that killing can be justified by appeal to associative duties. However, the Associativist Account should not prove too much: in particular, it would be a serious failing if could not justify something like the principle of noncombatant immunity, according to which belligerents in war ought to distinguish at all times between combatants and noncombatants, and direct their attacks only against the former. Combatants are members of the armed forces and others who directly participate in hostilities. Noncombatants are not combatants.

My objective is to argue that our associative duties can justify killing in war, without licensing violations of noncombatant immunity. Since we already know that these duties can override duties not to foreseeably but unintentionally kill others (whether combatants or noncombatants), the real challenge is to show that they can license some intentional killing of combatants, without also permitting intentional killing of noncombatants.⁶

In practice, the most important grounds for noncombatant immunity might be the contractualist and rule consequentialist arguments that the prohibition against killing noncombatants is the principal constraint on the calamities of war and can be sustained only if belligerent states trust one another to observe it. Given that warfare is unavoidably imbued with rights violations, it is imperative to minimise them insofar as is possible. On the contractualist interpretation, noncombatant immunity is the rule people, states, or decent states (depending on who is at the contractual table) would agree on, to best protect their own rights and interests against the suffering of war (Benbaji, 2008). For rule consequentialists, it is simply the best rule for

6. Note that on any plausible account, civilian political leaders who have genuine control over the military are in principle legitimate targets, excluded, if at all, only on the pragmatic grounds that killing them makes a lasting peace harder to secure.

all to observe, to reduce the calamities of war for all (Shue, 2010; Shue & Dill, 2012; Mavrodes, 1975; Waldron, 2010). The key difference between these two perspectives is that rule consequentialists affirm that this is the best rule overall, while contractualists claim that it is Pareto superior to alternatives. Of course, actually proving either of these empirical claims is impossible—not least because this is not a field where experimentation is permissible. Yet, the former claim is certainly very plausible, and while the latter might be a little more ambitious it might also still be reasonably affirmed. Sometimes belligerents might be unable to protect their own rights and interests without deliberately attacking noncombatants—but then, if their adversaries were not constrained by the principle of noncombatant immunity, their rights and interests would probably be more gravely threatened.

On these accounts, deliberately attacking noncombatants is not only wrongful because they retain their rights against attack (which is also often true of combatants), but it also undermines one's adversary's trust, which may lead to more rights-violating reprisals. It also undermines observance of the principle in the long term, as future belligerents will anticipate similar deviations. The obvious problem with this argument is that sometimes deliberately attacking noncombatants will in fact minimise the calamities of war and will not have these troubling consequences. In such cases contractualists and rule consequentialists struggle to explain why attacking noncombatants is impermissible.

To give the right answer in cases such as these, we need to stop thinking in terms of liability alone, and focus instead on degrees of wrongfulness. Even if some combatants are not liable to be killed, killing them is less wrongful than killing nonliable noncombatants. And even if combatants are justified in overriding some negative duties to protect their associates, they must still do as little wrong as they can—so they ought to target only combatants. I defend this view at length elsewhere (see Lazar, Forthcoming-b). Here, it must suffice to summarise my arguments.

First, as already noted, there is an important distinction between unavoidably harming a nonliable person in the course of eliminating a threat, and using him as a means to avert that threat. Attacks on combatants in war will often (though certainly not always) be eliminative: we kill them in order to avert the threat that they pose, or to which they contribute, or, if they do not threaten us themselves, in the course of killing other combatants who more directly threaten us. Attacks on noncombatants are, in this sense, often more opportunistic than attacks on combatants—the goal is to use the victims' suffering to break the society's political will. Combatants'

suffering is often not used in the same way. This distinction is not absolute—since noncombatants do contribute to their side's ability to fight, the agency involved in harming them can be partly eliminative; moreover combatants will often be killed opportunistically—used to break the will of their fellow combatants, or to create a diversion from another line of attack.

Second, it is more wrongful to intentionally kill nonliable people who you have greater reason to believe are not liable to be killed. While I argue at length elsewhere that not all combatants are liable to be killed, they are undoubtedly more likely to be liable than are noncombatants (Lazar, 2010a). And there are grounds to think that riskier killings of nonliable people are more wrongful, other things equal, than less risky killings. First, because the greater the likelihood that your victim is not liable to be killed, the more disrespect you show him by killing him nonetheless; second, because subjecting others to an unreasonable risk of suffering wrongful harm is an additional injury, over and above that of actually violating their right to life.

Third, attacking nonliable noncombatants may be more wrongful than attacking nonliable combatants because attacking combatants reflects their choices in a morally significant way. This idea is not new: Walzer, Benbaji and Hurka have argued that soldiers waive their rights against unjustified attack (Benbaji, 2008; Hurka, 2007; Walzer, 2006); McMahan contends that by voluntarily joining the armed forces, soldiers take a risk that can subsequently render them liable to be killed (McMahan, 2009, p. 183). My argument, however, is not that soldiers waive their rights to life. Instead, it is that killing combatants is less wrongful than killing noncombatants, simply because combatants have chosen to put themselves in the line of fire. Of course, if the combatant's only alternatives were so bad that his choice is not really voluntary, then this makes little difference—if the aggressor is genocidal, for example, then combatants' choice to fight clearly does not robustly reflect their will. In ordinary conflicts, however, the aggression is less murderous, and insofar as combatants have minimally acceptable alternatives to fighting, they do somewhat voluntarily expose themselves to risks, so harming them reflects their choices in a way that harming noncombatants does not. Moreover, most combatants fight to protect the lives and way of life of people whom they care about, so accordingly prefer that we engage them directly, rather than targeting the people they are trying to protect—though of course they would most prefer that we not fight at all.

The final plausible distinction between degrees of wrongdoing in war is at present the least studied. This is the notion that, other things equal, 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. Since noncombatants are usually more vulnerable than combatants, this renders killing nonliable noncombatants more wrongful than killing nonliable combatants.

The argument's descriptive claim is that noncombatants are normally more vulnerable to lethal attack than combatants. Obviously combatants are usually far better equipped than noncombatants, but they also enjoy other advantages—in particular, specialist training both as individuals and as a unit, which turns a group of individuals into a team, in which each plays a role in enhancing their collective security. Relatedly, combatants enjoy the normative resource of their comrades-in-arms' recognition of duties to protect their fellows.⁷ They know that their fellow combatants will take risks for their sake—and that they are trained to do so effectively. And combatants can often strike back—the most effective form of defence. Of course, they are sometimes also vulnerable—they cannot strike back against Predator drones, for example, and there is no defence against tactical nuclear weapons, though their training and coordination will still enable them to mitigate the dangers they face. But against a broad spectrum of threats combatants are clearly better able to defend themselves than are noncombatants. Indeed, if combatants were not less vulnerable than noncombatants, they would not be performing their function. Armed forces are supposed to coordinate and equip groups of people to attain strategic objectives through the use of force. This can be successful only if those people are better able than ordinary civilians to defend themselves and their units.

Why, though, should killing vulnerable, defenceless nonliable people be more wrongful than killing nonliable people who are not vulnerable or defenceless? First, because harming the defenceless marks the combatant out as cowardly and dishonourable. If combatants cannot fight morally pure wars, they are perpetually on the threshold between being warriors and murderers. The warrior ethos enables them to assert their position on the right side of that line. Attacking the defenceless is a violation of this ethos—it is shameful cowardice.

Second, radical vulnerabilities create responsibilities (Goodin, 1985). If and insofar as A is especially vulnerable to B, B has some responsibilities to A. The dramatic power difference between them generates duties. This is one explanation of why adults have duties to protect children, and why societies have a duty to look after the weakest and most vulnerable. Noncombatants are radically vulnerable to com-

7. Thanks to Nancy Sherman for this point.

batants, who hold their lives in their hands. Attacks on noncombatants exploit this power differential, betraying the duty of care owed by the powerful to the weak. The flip side of this is that the threat of death for a nonliable noncombatant is profoundly disempowering: their vulnerability denies them control over their fate, as they are subject to an arbitrary and lethal will. This is a further reason against targeting the vulnerable.

These are just outlines of arguments that I make in greater depth elsewhere (Lazar, Forthcoming-b). But the overall point should be clear: noncombatant immunity does not depend on affirming that noncombatants retain rights that combatants lose. That the Associativist Account can justify overriding rights, then, does not necessarily mean it will legitimate attacks on noncombatants. Rights violations can be more and less wrongful. When we use a person's suffering as a means to compel others, when we shoot knowing there is a high chance the target is not liable and when our victim is defenceless, we are committing especially egregious breaches of negative duty, and these cannot be justified by appeal to our associative duties. By contrast, when we have more reason to believe our target is liable, when we are not using his suffering as a means, when he is able to defend himself, and he has chosen to put himself in the firing line, then the breach of negative duty, though still very grave, is less wrongful, and can more readily be overridden.

As an aside: one might think that the asymmetrist account could deploy these various defences of noncombatant immunity, in response to my earlier critique of its deleterious impact on the protection of noncombatants in war (e.g. Lazar 2010a). That critique, however, focused on the low threshold version of the asymmetrist position, according to which, in my view, too many noncombatants are liable to be killed in war. The foregoing arguments are considerably less plausible on a low threshold interpretation than they are on a high-threshold view. If noncombatants were liable to be attacked, then we would be less concerned to reduce harms to them, so the rationale for the rule-consequentialist and contractualist justifications of noncombatant immunity would break down. The eliminative/opportunistic agency distinction is little use to the low threshold view, since if you are liable to be harmed eliminatively, then you are almost certainly liable to be harmed opportunistically to the same degree. On a low threshold account, at least in many contemporary wars, enough adult noncombatants are going to be liable to be killed that, in targeting them, you do not unduly risk harms to nonliable people. Whether or not killing you reflects your will is, again, much less relevant when you are liable to be killed anyway. Similarly, if

you are liable to be attacked, then whether you are defenceless or vulnerable appears irrelevant.

RESTRICTING THE ASSOCIATIVE DUTY TO PROTECT

One might object against the Associativist Account that the reasons it draws on are detached from regular morality, offering a distinct source of normativity potentially radically at odds with our standard moral reasons, which might have very troubling implications outside of war. There are three important variants of this objection. The first notes that sometimes our associates are liable to be harmed—sometimes they even deserve it. Does the Associativist Account ground duties to protect them in such cases? The second variant notes that sometimes really evil people can develop deep relationships with one another. Suppose notorious serial killers Myra Hindley and Ian Brady had an (internally) morally valuable relationship. Would they have had associative duties to one another? The third asks whether granting associative duties this degree of gravity might entail troubling conclusions about what ordinary people can do to protect our associates outside of war.

The first response to the first objection is that the Associativist Account is probably not committed either way—this is a question for one's full theory of associative duties. Still, we can consider how the argument might go, and it does seem clear that we have weaker duties to protect our associates, the more liable they are. Suppose A can defend his associate B from a threat posed by C, which is in response to a threat B posed to C's life. If B is at fault for the threat she poses, then A's duties to protect her are much weakened, and certainly cannot override his duties not to harm C. Notice, though, that B's liability does not defeat all A's duties to protect her. If he could save B without C being harmed, A will have a duty to protect B. Suppose, for example, that A can save B by interposing his body between B and C, taking the force of C's blow. For some magnitudes of resultant harm, and some relationships, A will have an associative duty to B to bear that harm to protect her, but would not have a general duty to do so because his general duties to protect others are less stringent than his associative duties to do so.

By contrast, it seems unlikely that we have duties to protect associates from harms they deserve to suffer. Suppose B has murdered someone, and has fled the police. She comes to A seeking shelter. Though A might want to protect her from the threat of capture, he surely has no morally grounded reason to do so. B deserves

to be caught and punished, and any duties A has to protect her are defeated by that fact. That noted, he may still be justified in recusing himself from the pursuit of his associate: though he has no duty to protect her from deserved threats, he perhaps cannot be required to be the agent of retribution. It will matter, though, whether he has other reasons to be that agent. In Bruce Springsteen's song 'Highway Patrolman', from the album *Nebraska*, he describes how a police officer, Joe Roberts, allows his brother (named Frankie—Frankie ain't no good) to escape to Canada after a fight in which he murdered another man. He follows Frankie's car across town, then as they get close to the Canadian border stops giving chase.⁸ This seems an understandable but unjustifiable exaggeration of associativist reasons—the brother deserves to be caught, tried and punished, so Joe has no duty to protect him from that fate. Given his duties as a police officer, Joe may not recuse himself in this way—unless other officers will catch Frankie without his help, which is not the case. However, if he were not a police officer, I think he would be permitted to allow his brother to escape.

Our duties to protect our associates from harms to which they are liable cannot weigh against our general duties to any significant degree. If they deserve to be harmed, we may have no duty to protect them at all. These are important constraints on the role of associative duties in morality. However, they do not undermine the Associativist Account of killing in war—provided that we combine it with a particular view of what grounds liability to be killed in war. Theories of liability typically argue that an individual must have made some sort of contribution to an unjustified threat to be liable to be killed, when that is a necessary and proportionate means to avert the threat. They vary depending on what degree of contribution to the threat they think liability presupposes. On low threshold views, a small causal contribution, for which one is minimally responsible, can be sufficient for liability to be killed. On high threshold views, a much more significant causal contribution is required, or some significant degree of culpability. On low threshold views, many combatants and noncombatants in war (at least on the unjustified side, if there is only one) will be liable to be killed. On the high threshold view—which I affirm—almost all noncombatants will not be liable to be killed, indeed nor will a significant proportion of combatants. It follows that at least some of the associative duties owed by comrades-in-arms to protect one another, and almost all the associative duties owed by noncombatants to protect one another, will be owed to people who are not liable to be killed.

The second worry derives from the possibility that people committed to deeply

8. Thanks to David Rodin for this example.

evil activities might develop deep relationships with one another. If these relationships ground duties through their internal qualities, then the disconnect between associativist reasons and ordinary morality becomes too great. A powerful example is reserve police battalion 101, called up during the second world war by the Nazis, and tasked with exterminating Polish Jews (Browning, 2001). Members of this battalion reported strong feelings of loyalty to one another, which in part motivated their participation in the atrocities. Similar stories are reported from American atrocities in Vietnam, and are dramatised in films like *Casualties of War*. Does the Associativist Account open the door to this sort of deeply objectionable reasoning?

The first response is that reasons of loyalty were in these cases clearly being abused. Members of battalion 101 were not protecting one another against a threat. Their victims were unarmed men, women and children. An appeal to associative duties is as specious as appealing to self-defence in these cases.

Second, they were liable to whatever threats they did face, and indeed probably deserved them. One cannot nonculpably engage in genocide. Thus any associative duties they might have had to one another would have been either defeated or seriously weakened, as per the foregoing argument.

A third response more directly questions whether relationships that are predicated on doing evil to outsiders can really generate associative duties. If the relationship's internal qualities were developed through the knowing infliction of wrongful suffering on outsiders, then any capacity that relationship would otherwise have had to generate strong moral reasons may well be defeated (for a similar argument see Hurka, 1997). Suppose the members of reserve police battalion 101 went through a lot together, as they murdered thousands of Polish Jews. Perhaps they sacrificed themselves for one another, risking their lives to protect each other against the few people among their victims who still posed a threat. Any property of the relationship that traces its history to the knowing infliction of such egregiously wrongful suffering loses any reason-giving capacity it might otherwise have had.

Given these arguments, we can see where the Associativist Account stands on the debate between Walzerian symmetrists and anti-Walzerian asymmetrists, mentioned in the introduction. Evidently the Associativist Account starts with the belief that neither Walzer nor his critics can justify fighting in wars where we must intentionally kill nonliable people. Since it endorses the high threshold view of liability to be killed, it asserts that in all likely wars victory depends on intentionally violating some people's rights to life, so we need additional moral reasons to override those

rights violations, some of which we can find in associativist morality. In that sense, it offers a different perspective on the morality of war from both established camps.

However, underlying the specific terms of the debate between Walzer and his critics is a broader question, on which the Associativist Account does take sides. For Walzer, soldiers need not ask themselves whether fighting is justified. They know that if they adhere to the standards for just conduct in war, they will fight justly, without violating rights. Thus it is always permissible to participate in wars. For Walzer's critics, this view is both pernicious and inane. As Christopher Kutz puts it: 'if death and destruction matter morally, as they do, and if reasons matter morally, as they do, then differences in combatants' reasons for bringing about death and destruction must also matter morally' (Kutz, 2008, p. 44). The Associativist Account agrees with Walzer's critics: inflicting death and destruction does matter morally, so soldiers may fight only if they have very strong reasons for doing so. If they do not have sufficient reason to fight and kill, then they are acting unjustifiably, even if they adhere to the *in bello* code.

However, the Associativist Account joins Walzer in affirming that combatants on both sides of a war can, in some cases, fight justifiably. Most importantly, it explains how combatants-J can justifiably fight, even though doing so inevitably involves violating rights. Since the primary goal of any theory of the ethics of war that denies pacifism should be to show that soldiers on at least one side in at least some wars can fight justifiably, this is the headline result.

But the Associativist Account can also give combatants-U reason to fight and kill, insofar as they are performing their own associative duties to protect their comrades-in-arms from wrongful harms, and acting on behalf of a community whose members have duties to protect their associates from wrongful harms (i.e. those the victim is not liable to suffer). Of course, sometimes the best way for combatants-U to protect their associates will be to surrender. I do not mean to exaggerate the justificatory force of the Associativist Account, or the merits of warfare as a means of protecting those we care about. Moreover, combatants-U clearly face a heavy burden of justification, insofar as their fighting contributes to their side's achieving its unjustified goals—this will probably mean that associative duties could only justify genuinely defensive operations, for example. Nonetheless, *contra* Walzer's critics, combatants-U can sometimes be justified in fighting, and on much the same grounds as combatants-J.

The final worry is that the Associativist Account will justify egregiously wrong-

ful partiality outside of war. For example, suppose A's associate, B, is dying from organ failure, and the only way to save her is to kill C and transplant his organs—A's will not work, for medical reasons. If duties to protect our associates can override general negative duties not to kill innocent people, then why should A's duty to protect B not override his duty not to harm C? Indeed, given the argument about institutionalisation advanced in the last section, couldn't we argue that we are justified, as a society, in setting up institutions where doctors routinely harvest organs from some, so that we can perform our associative duties to protect our loved ones from organ failure? This would obviously be an outrageous result.

Fortunately, however, it is not a conclusion that the Associativist Account needs to draw, or even one that it plausibly licenses. Killing C to transplant his organs is an egregiously wrongful form of killing, and A's duty to protect B does not override his duty not to kill C in this way. 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. I argued further that there are other respects in which the killing necessary to win wars is less wrongful than other forms of killing—there is some chance that the combatants we kill are liable, killing them somewhat reflects their choices, they are less vulnerable and defenceless than noncombatants, and so on. None of these considerations applies here. Killing C is egregiously opportunistic, he is clearly not liable to be killed, he is vulnerable and defenceless, and harming him in no way reflects his choices. Since A would not be permitted to kill C in this way to save B, there are no grounds for applying the reasoning of this paper to justify institutionalised organ-harvesting or other similarly counterintuitive schemes.

CONCLUSION

The Associativist Account is not a complete theory of the morality of war. In particular, it cannot justify the resort to war in defence against political aggression, and it cannot justify the intentional, opportunistic killing of nonliable people that warfare inevitably involves. However, it can justify some of the intentional killing in war. The foregoing discussion is intended to dispel the instinctive rejection of the Associativist Account by philosophers who are predictably sceptical about the deployment of reasons of loyalty in moral theory. It has offered grounds for those duties and argued, against the philosophical orthodoxy, that associative duties can override

serious general negative duties, both in peacetime and in war. It has also shown that conceding this point does not mean allowing a tribalist anti-egalitarianism to run amok through morality. It is quite possible to concede the force of associative duties in war while still restraining their impact, both in war and in ordinary life.

Acknowledgements: I first developed these ideas in my D.Phil. thesis, War and Associative Duties (2009). My greatest debts are to Henry Shue, Jeff McMahan, and David Rodin, as well as other colleagues in the Oxford war workshop: Cécile Fabre, Per Ilsaas, Zahler Bryan, and Janina Dill. I presented versions of the paper at Oxford's Uehiro Centre, its Centre for Ethics, Law and Armed Conflict, and the Nuffield political theory workshop, as well as to audiences in Dartmouth, Washington University St Louis, Chicago, Toronto, Stanford and Melbourne. Thanks to all those who participated in those events, and to their organisers for inviting me. Particular thanks to Michael Bratman, Kimberley Brownlee, Josh Cohen, Julia Driver, Tom Hurka, John Filling, David Miller, Julian Savulescu, Andrew Sepielli, Nic Southwood, Adam Swift, Laura Valentini, Juri Viehoff, and Lea Ypi. Thanks also to an anonymous reviewer for this Journal, whose comments were particularly helpful. And thanks to Lachlan Umbers for research assistance. Research on this project was supported by an Australian Research Council Discovery Early Career Researcher Award.

REFERENCES

- Altman, A. and C.H. Wellman. 2008. 'From Humanitarian Intervention to Assassination: Human Rights and Political Violence'. *Ethics* 118/1: 228-257.
- Ashford, E. 2003. 'Individual Responsibility and Global Consequences'. *Philosophical Books* 44/2: 100-10.
- Barnham, D. 1975. *One Man's Window*, London, New English Library.
- Benbaji, Y. 2008. 'A Defense of the Traditional War Convention'. *Ethics* 118/3: 464-95.
- Bourke, J. 1999. *An Intimate History of Killing: Face-to-Face Killing in Twentieth-Century Warfare*, London, Granta.
- Brighouse H. and A. Swift. 2006. 'Parents' Rights and the Value of the Family'. *Ethics* 117/1: 80-108.
- Brighouse, H. and A. Swift. 2009. 'Legitimate Parental Partiality'. *Philosophy & Public Affairs* 37/1: 43-80.
- Browning, C.R. 2001. *Ordinary Men: Reserve Police Battalion 101 and the Final Solution in Poland*, London, Penguin.
- Coady, C.A.J. 2008. *Morality and Political Violence*, Cambridge, Cambridge University Press.

- Cottingham, J. 1983. 'Ethics and Impartiality'. *Philosophical Studies* 43/1: 83-99.
- 1986. 'Partiality, Favouritism and Morality'. *The Philosophical Quarterly* 36/144: 357-73.
- Dancy, J. 1993. *Moral Reasons*, Oxford, Blackwell.
- 2004. *Ethics without Principles*, Oxford, Clarendon Press.
- Emerton, P. and T. Handfield. 2009. 'Order and Affray: Defensive Privileges in Warfare'. *Philosophy & Public Affairs* 37/4: 382-414.
- Etzioni, A. 2002. 'Are Particularistic Obligations Justified? A Communitarian Examination'. *The Review of Politics* 64/4: 573-98.
- Fabre, C. 2009a. 'Guns, Food, and Liability to Attack in War'. *Ethics* 120/1: 36-63.
- 2009b. 'Permissible Rescue Killings'. *Proceedings of the Aristotelian Society* 109: 149-64.
- Friedman, M. 1991. 'The Practice of Partiality'. *Ethics* 101/4: 818-35.
- Frowe, H. 2008. 'Equating Innocent Threats and Bystanders'. *Journal of Applied Philosophy* 25/4: 277-90.
- Goodin, R.E. 1985. *Protecting the Vulnerable: A Reanalysis of our Social Responsibilities*, Chicago, University of Chicago Press.
- Gray, J.G. 1998. *The Warriors: Reflections on Men in Battle*, London: University of Nebraska Press.
- Hardimon, M.O. 1994. 'Role Obligations'. *The Journal of Philosophy* 91/7: 333-63.
- Horton, J. 2006. 'In Defence of Associative Political Obligations: Part One'. *Political Studies* 54/3: 427-43.
- 2007. 'In Defence of Associative Political Obligations: Part Two'. *Political Studies* 55/1: 1-19.
- Hurka, T. 1997. 'The Justification of National Partiality'. In McKim, R. and McMahan, J. eds. *The Morality of Nationalism*, Oxford, Oxford University Press, 139-57.
- 2005. 'Proportionality in the Morality of War'. *Philosophy & Public Affairs* 33/1: 34-66.
- 2007. 'Liability and Just Cause'. *Ethics & International Affairs* 21/2: 199-218.
- Jeske, D. 1996. 'Associative Obligations, Voluntarism, and Equality'. *Pacific Philosophical Quarterly* 77/4: 289-309.
- Kamm, F.M. 1996. *Morality, Mortality: Volume II*, Oxford, Oxford University Press.
- 2004. 'Failures of Just War Theory: Terror, Harm, and Justice'. *Ethics* 114/4: 650-92.
- Keller, S. 2006. 'Four Theories of Filial Duty'. *The Philosophical Quarterly* 56/223: 254-274.
- *The Limits of Loyalty*, Cambridge: Cambridge University Press.
- Kolodny, N. 2002. 'Do Associative Duties Matter?'. *Journal of Political Philosophy* 10/3: 250-66.
- Kutz, C. 2008. 'Fearful Symmetry'. In Rodin, D. and Shue, H. eds. *Just and Unjust Warriors: The Moral and Legal Status of Soldiers*, Oxford, Oxford University Press, 69-86.
- Lazar, S. 2009a. 'Responsibility, Risk and Killing in Self-Defence'. *Ethics* 119/4L: 699-728.
- 2009b. 'The Nature and Disvalue of Injury'. *Res Publica* 15/3: 289-304.

- 2009c. 'Do Associative Duties Really Not Matter?'. *Journal of Political Philosophy* 17/1: 90-101.
- 2010a. 'The Responsibility Dilemma for Killing in War'. *Philosophy & Public Affairs* 38/2: 180-213.
- 2010b. 'A Liberal Defence of (Some) Duties to Compatriots'. *Journal of Applied Philosophy* 27/3: 246-57.
- Forthcoming-a. 'National Defence, Self-Defence, and the Problem of Political Aggression'. In Lazar, S. and Fabre, C. eds. *The Morality of Defensive War*, Oxford, Oxford University Press.
- Forthcoming-b. *The Principle of Distinction between Combatants and Noncombatants: A Defence*, Oxford, Oxford University Press.
- Unpublished Manuscript. 'On The Justification of Associative Duties'.
- Lee, D. 2006. *Up Close and Personal: The Reality of Close-Quarters Fighting in World War II*, London, Greenhill.
- Lefkowitz, D. 2009. 'Partiality and Weighing Harm to Non-Combatants'. *Journal of Moral Philosophy* 6/3: 298-316.
- Mason, A. 2000. *Community, Solidarity and Belonging: Levels of Community and Their Normative Significance*, Cambridge, Cambridge University Press.
- Mavrodes, G. I. 1975. 'Conventions and the Morality of War'. *Philosophy and Public Affairs* 4/2: 117-31.
- McMahan, J. 1994. 'Innocence, Self-Defense and Killing in War'. *Journal of Political Philosophy* 2/3: 193-221.
- 1997. 'The Limits of National Partiality'. In McKim, R. and McMahan, J. eds. *The Morality of Nationalism*, Oxford, Oxford University Press, 107-138.
- 2004. 'The Ethics of Killing in War'. *Ethics* 114/1: 693-732.
- 2009. *Killing in War*, Oxford, Oxford University Press.
- Forthcoming. 'What Rights May Be Defended by Means of War?'. In Lazar, S. and Fabre, C. eds. *The Morality of Defensive War*, Oxford, Oxford University Press.
- McPherson, L. 2004. 'Innocence and Responsibility in War'. *Canadian Journal of Philosophy* 34/4: 485-506.
- Miller D. 2005. 'Reasonable Partiality Towards Compatriots'. *Ethical Theory and Moral Practice* 8/1: 63-81.
- Miller, S. 2007. 'Civilian Immunity, Forcing the Choice, and Collective Responsibility'. In Primoratz, I. ed. *Civilian Immunity in War*, Oxford, Oxford University Press, 113-35.
- Nicol, M. 2007. *Condor Blues: British Soldiers at War*, Edinburgh, Mainstream.
- Øverland, G. 2006. 'Killing Soldiers'. *Ethics & International Affairs* 20/4: 455-75.

- Ninh, B. 1993. *The Sorrow of War: A Novel*, London, Minerva.
- O'Brien, T. 1991. *The Things They Carried*, London, Flamingo.
- Owen, W. 1996a. 'Anthem for Doomed Youth'. In Silkin, J. ed. 1996. *The Penguin Book of First World War Poetry*, London, Penguin, 193-4.
- Owen, W. 1996b. 'Apologia Pro Poemate Meo'. In Silkin, J. ed. *The Penguin Book of First World War Poetry*, London, Penguin, 197-8.
- Pogge, T. 2002. 'Cosmopolitanism: A Defense'. *Critical Review of International Social and Political Philosophy* 5/3: 86-91.
- Quinn, W.S. 1989. 'Actions, Intentions, and Consequences: The Doctrine of Double Effect'. *Philosophy and Public Affairs* 18/4: 334-351.
- Quong, J. 2009. 'Killing in Self-Defense'. *Ethics* 119/2: 507-37.
- Raz, J. 1986. *The Morality of Freedom*, Oxford, Clarendon Press.
- 1989. 'Liberating Duties'. *Law and Philosophy* 8/1: 3-21.
- Remarque, E.M. 1929. *All Quiet on the Western Front*. Translated by B. Murdoch., 1996. London: Vintage.
- Norton-Taylor, R. 2006. 'Bullying Systematic at Deepcut, Says QC's Report'. *The Guardian*. Retrieved from: <http://www.guardian.co.uk:uk:2006:mar:30:military.richardnortontaylor>.
- Roberts, A. 2010. 'Lives and Statistics: Are 90% of War Victims Civilians?'. *Survival* 52/3: 115-136.
- Scanlon, T.M. 1998. *What We Owe to Each Other*, London, Belknap.
- Scheffler, S. 2002. *Boundaries and Allegiances: Problems of Justice and Responsibility in Liberal Thought*, Oxford: Oxford University Press.
- 2010. 'Morality and Reasonable Partiality'. In Feltham, B. and Cottingham, J. eds. *Partiality and Impartiality*, Oxford, Oxford University Press, 98-130.
- Seglow, J. 2010. 'Associative Duties and Global Justice'. *Journal of Moral Philosophy* 7/1: 54-73.
- Shue H. and J. Dill. 2012. 'Limiting the Killing in War: Military Necessity and the St. Petersburg Assumption'. *Ethics and International Affairs* 26/3: 311-33.
- Shue, H. 2008. 'Do We Need a Morality of War?'. In Rodin, D. and Shue, H. eds. *Just and Unjust Warriors: The Moral and Legal Status of Soldiers*, Oxford, Oxford University Press, 2008, 87-111.
- 2010. 'Laws of War'. In Besson, S. and Tasioulas, J. eds. *The Philosophy of International Law*, New York, Oxford University Press, 511-30.
- Slim, H. 2007. *Killing Civilians*, Oxford, Oxford University Press.
- Steinhoff, U. 2008. 'Jeff McMahan on the Moral Inequality of Combatants'. *Journal of Political Philosophy* 16/2: 220-6.
- Tadros, V. 2012. *The Ends of Harm: The Moral Foundations of the Criminal Law*, Oxford, Oxford University Press.

Tamir, Y. 1993. *Liberal Nationalism*, Princeton, Princeton University Press.

Temkin, L. 2012. *Rethinking the Good*, Oxford, Clarendon Press.

Waldron, J. 2010. *Torture, Terror, and Trade-Offs: Philosophy for the White House*, Oxford, Oxford University Press.

Walzer, M. 2006. *Just and Unjust Wars: A Moral Argument with Historical Illustrations*, 4th ed., New York, Basic Books.

Wolf, S. 1992. 'Morality and Partiality'. *Philosophical Perspectives* 6: 243-259.

Biotechnology, Justice and Health

RUTH FADEN

Johns Hopkins Berman Institute of Bioethics

MADISON POWERS

*Department of Philosophy and Kennedy Institute of Ethics,
Georgetown University*

ABSTRACT

New biotechnologies have the potential to both dramatically improve human well-being and dramatically widen inequalities in well-being. This paper addresses a question that lies squarely on the fault line of these two claims: When as a matter of justice are societies obligated to include a new biotechnology in a national healthcare system? This question is approached from the standpoint of a twin aim theory of justice, in which social structures, including nation-states, have double-barreled theoretical objectives with regard to human well-being. The first aim is to achieve a sufficient level of well-being in each of six core dimensions. In the special case of healthcare systems, this aim is focally but not exclusively attentive to achieving health sufficiency as one of the core dimensions. The second aim is to combat the emergence and persistence of densely woven patterns of systematic disadvantage that tend to undermine the achievement of a sufficient level of health and the other core elements of well-being of some persons and groups. Judgments about entitlements to health related resources, including new biotechnologies, are made in light of a threshold notion of health sufficiency. What is enough or sufficient health? The answer that is defended here is that sufficient health is enough health for a decent human life, understood as enough health to live a full life course without preventable, significant functional disability or decrement in health, or treatable pain or suffering. When a state must include a new biotechnology in its national healthcare system is also influenced by ancillary concerns about the connection between health and other core dimensions of well-being. What counts as a significant functional impairment or health decrement is thus explicated, in part, in relation to the theory's sufficiency aim for the

other essential dimensions of well-being, and thus for a decent life, overall. Those elements of health that play a critical role in the experience of sufficient reasoning, affiliation, security, respect and self determination are especially important; any loss of health function or capacity that threatens the individual's prospects for sufficiency in these other dimensions, including the relational egalitarian concerns they entail, constitutes a significant functional impairment. Within national borders, individuals are thus entitled to those health-related goods and services that are essential for a sufficiency of each of the dimensions of well-being; with regard to self determination and respect, what is sufficient by way of guaranteed access to specific goods and services is going to depend on the implications of such access for where an individual stands in relation to her co-nationals. The content of any entitlement to health-related goods and services is also necessarily dynamic. What can be done for health and the other core dimensions of well-being as a function of technological innovation and diffusion is in constant flux. The paper concludes by considering the implications of this analysis for the conditions under which states are obligated to include access in their healthcare systems to one biotechnology, deep brain stimulation.

INTRODUCTION

Deep brain stimulation, or DBS, is a biotechnology in which guide wires are inserted through the skull and into the brain, where they deliver electrical currents to clusters of neurons that are no longer performing properly. DBS was first employed to control the tremors of Parkinson's disease and is now routinely used in many countries to treat other movement disorders as well, such as seizures and Tourette's, and also increasingly recalcitrant depression.

In early 2012, the *New England Journal of Medicine* reported the results of an experiment to use DBS for yet a different purpose- to boost failing memory (Suthana et al, 2012). Two key memory regions of the brain were stimulated in 7 seizure patients, all of whom showed improved cognitive performance during stimulation. These findings have generated excitement about the prospect that DBS might become an effective treatment for dreaded memory disorders like Alzheimer's and also about

the prospect that DBS might someday be used to enhance memory in those with no functional memory loss.

New biotechnologies like DBS have the potential to both dramatically improve human well-being and dramatically widen inequalities in well-being. In this paper, we tackle a question that lies squarely on the fault line of these two claims.

When, as a matter of justice, are societies obligated to include a new biotechnology in a national health care system or otherwise assure that all have reasonable access to it?

The paper proceeds in three parts. We begin by explicating one way of thinking about this question by providing a very brief sketch of our twin aim theory of justice as presented in our book, *Social Justice* (Powers & Faden, 2006), and a new book currently in progress. In Part Two, we focus on the concepts of a decent life and sufficiency, which play prominent roles in our theory in helping to explicate what goods and services individuals are entitled to receive. In Part Three, we illustrate some of the implications of our approach by looking more carefully at the example of DBS in these promising new memory applications.

PART ONE: A TWIN AIM THEORY OF JUSTICE

For us, claims in justice to healthcare goods and services draw their foundational legitimacy—their ultimate justificatory structure—from the essential and direct role that these goods play in human well-being, which we take to be the primary object of social justice. Ours is a twin-aim theory of justice, in which social structures, including both global institutions and nation-states, have double barreled objectives or aims with regard to human well-being. The target of each barrel is morally distinct from the other, but strategically reinforcing.

One aim is the improvement of well-being. In the special case of healthcare systems, this aim is focally but not exclusively attentive to improving health as an essential dimension of well-being. The other aim is to combat, whether by prevention or amelioration, the adverse effects on well-being caused by densely woven patterns of systematic disadvantage that profoundly compromise or will compromise the health and other core elements of well-being of some persons and groups.

The well-being aim begins with a broadly Aristotelian-inspired effort to identify some core components or elements of human well-being that are among the suitable objects of concern within a theory of justice. We specify core elements of human well-

being as having three characteristics. They are (1) typical of normal human development and valuable for their own sakes, apart from any further good they might bring; (2) of value to anyone, on reflection, whatever their particular life projects, chosen activities, or personal aspirations might be; and (3) humanly alterable and profoundly and pervasively influenced in their development and maintenance by basic social institutions and social practices.

THE WELL-BEING AIM AND ITS CORE ELEMENTS

The well-being aim of social justice is defined in terms of a level of sufficiency of each of its core elements. Although it is beyond our task here to fully engage the thorny questions of just how sufficiency for any element is assessed, we will have more to say about this later, particularly with regard to health. The key point for now is that sufficiency represents a moral minimum of justice. Inequalities in well-being in which individuals fail to meet even this moral minimum are among those most morally urgent to address.

Our list includes six core elements. Though these six overlap in terms of both determinants and effects, each is of independent and equal moral value.

1. We have already mentioned health. Although health is the element of human development and flourishing most intimately connected with the biological or organic functioning of the body, the absence of health refers to more than biological malfunctioning or impairments. Being in pain, even if that pain does not impede proper biological functioning; sexual dysfunction; and infertility are also incompatible with health.

2. The second element is personal security from actual physical and psychological harm as well, as the threat of such harm.

3. The third element is the development of reasoning, both for deliberation and choice and for the formation of beliefs and inferences regarding facts about the natural and social world.

4. The fourth element involves the formation and maintenance of personal at-

tachments of various kinds with varying degrees of intimacy, commitment, and affective engagement.

5. The fifth element is the respect of others. Here, the central concern is that individuals be able to live under social conditions in which others judge and treat them as moral equals, as persons worthy of the same sort of treatment any other person merits, and among co-nationals, as fellow citizens, as persons having comparable political standing.

6. The sixth element is self-determination, the ability to shape the broad contours of a life, to have some significant say over the general course of one's life. Put slightly differently, self-determination is about living a life that is not in crucial respects under the domination and control of others or the tyranny of profound necessity.

The importance of these last two elements to our overall account of well being cannot be overstated. Whether in the domestic case, in which equal moral standing and political standing are at issue, or in the international case, where moral equality is the paramount concern, the intent is that none is either at grave risk for the tyranny of deprivation or the threat of domination or oppression of others. Thus, among the defining concerns expressed in the positive aim of ensuring a sufficiency of well-being are two relational dimensions of well-being -- the good of living a self-determining life and the good of having a social standing involving the respect of others— where sufficiency requires nothing less than equality itself.

THE SYSTEMATIC DISADVANTAGE AIM

While the first aim focuses on the injustice of failures in the realization of a threshold level of the core elements of well-being, the second aim focuses on the injustice of creating and maintaining serious social impediments to sufficient levels of well-being. These impediments are wrong on our account even if some systematically disadvantaged individuals and members of disadvantaged groups manage to escape the adverse consequences to their well-being that most others so disadvantaged predictably experience.

PART TWO: SUFFICIENCY AND A DECENT HUMAN LIFE

Put another way, the first aim locates injustice in certain *adverse consequences* for well-being, while the second aim locates injustice in the *structural unfairness* of forms of social arrangement that distribute advantages and disadvantages in particular ways.

There is much that could be said about the work that the notion of a decent human life does within our theory. A few comments should suffice for now. A decent life requires more than mere subsistence, for example, but less than a maximally flourishing life for all. Indeed, these very points are familiar ones from within contemporary discussions of human rights where the aim is to articulate the minimum demands of a transnational standard of well-being that is neither implausibly high in what it would require of those upon whom such a duty might fall nor so low that it demands little more than bare subsistence (Nickel, 2007, 61-66, 98-103, 138-42; Buchanan, 2010).

In addition, our account of a decent life requires satisfaction of the moral demands of the relational egalitarian concerns expressed by a commitment to the goods of respect and self-determination, an emphasis which is also congenial with one strain in the human rights literature that expresses similar relational concerns. James Nickel, for example, suggests that there is a place in human rights for a variety of reasonable claims other than those for food, shelter, and education, including concerns for some measure of control over one's own life and some social guarantees of equal standing in political and legal arenas (Nickel, 2007, 63-64).

With the emphasis on self-determination and respect in our account of well being firmly in place, we can make a few observations about what would be *contrary* to a decent life. A decent life is not one that is marked by servility, slavishness, the necessity to grovel, or deep dependence on the good will or whim of others for the most basic requirements of life. A decent life is not one marred by the most degrading aspects of poverty, including squalor, helplessness, extreme vulnerability, or the inability to provide for one's own children and family (Powers & Faden, 2006, 138-41; Margalit, 1998). A decent life is not one in which someone is treated as less than a full member of the human community or as someone whose most vital concerns are accorded little or no weight. Indeed, to be treated as a second-class citizen within one's own society is to be treated as a second-class human being, as someone who is not due the full measure of respect and concern accorded to other members of a common political community (Margalit, 1998).

A decent human life is one in which someone is not subjugated, marginalized, stigmatized, infantilized, or deprived of the full use of one's mature faculties.¹² Much more might be said along these lines, but the essential point we make here is that these relational aspects of well-being earn a place on the list of constituents of a decent human life as securely as concerns about health, cognitive development, and physical security. Moreover, they function as a check on how we should understand what counts as sufficiency for other constituents such as health.

Both the notion of a decent human life and the relational egalitarian concerns that are built into our conception of well-being sufficiency bear on what counts as sufficiency in health. They also bear on the answers we give to questions of entitlements to health care goods and services instrumental to health sufficiency. We judge what is enough in the realm of health, for example, in part by reference to the impact of relative inequalities on social standing and the ability of individuals and groups to lead self-determining lives and avoid the kinds of conditions inimical to a decent human life. We then make judgments of entitlements to health related resources, *including new biotechnologies*, in light of that notion of health sufficiency, influenced by ancillary concerns about the connection between health and other core dimensions of well-being and how resource policy choices impact the full set of well-being concerns.

HEALTH SUFFICIENCY AND HEALTHCARE ENTITLEMENTS

What is enough or sufficient health? And what instrumental goods and services, what resources, are individuals entitled to as a consequence?

Our answer to the first question is-- enough health for a decent human life. For us, this means enough health to live a full life course without significant disability or decrement in health, or pain or suffering that can be reasonably prevented, treated or relieved.

Clearly and intentionally, this is a higher threshold than one that would require only enough health to escape severe pain or suffering, death in childhood or early adulthood, or the most egregious of disabilities. Any account of a decent human life must include living through all the stages of life without significant disability or dec-

1. Many of these points are developed in Mill's *The Subjection of Women*, which for us has been influential in our own thinking about what is central to well-being in view of a decent human life.

rement in health that can be reasonably prevented or mitigated, and pain or suffering that can reasonably be relieved.

At the same time, however, we are also far from what would be demanded for a maximally flourishing life, which would translate to a threshold of something like maximal health for all. Merely because the world's oldest person just celebrated his 115th birthday in Japan, or even because increasing numbers of people in many countries are now living into their 90s, it does not follow that the threshold of health sufficiency requires all to do so. It is sufficient that all live through each stage of life, including old age. Similarly, health sufficiency does not require the absence of any and all limitations on health-related quality life. It is enough that all live lives unmarked by significant disabilities or decrements in health that could reasonably be prevented or mitigated, and pain and suffering that could be alleviated.

Importantly, what counts for us as a significant health decrement is explicated, in part, in relation to our theory's sufficiency aim for the other core dimensions of well-being, and thus for a decent life, overall. Those elements of health that play a critical role in the experience of sufficient reasoning, affiliation, security, respect and self determination are especially important; any loss of health function or capacity that threatens the individual's prospects for sufficiency in any of the other dimensions, including the relational egalitarian concerns they entail, constitutes a significant impairment.

What instrumental goods and services, what resources, are individuals entitled to as a consequence of this sufficiency threshold?

Our formal answer is deceptively simple: Individuals are entitled to those health-related goods and services that are required for a sufficiency of health and the other essential dimensions of well-being, as needed for a decent life. With regard to self determination and respect, what is sufficient by way of guaranteed access to specific goods and services is going to depend on the implications of such access for where an individual stands in relation to her co-nationals.

It is implausible to think that the content of any entitlement to health-related goods and services, global or domestic, would be fixed for all time. Rather, the content is necessarily dynamic. What can be done to protect, restore and promote health and the other essential dimensions of well-being, is, as a function of technological innovation and diffusion, as well as cost and value, in constant flux.

At minimum, interventions that produce at least moderate value at no more than moderate cost are likely candidates for inclusion in the content of the entitle-

ment. For us, the value of a health intervention is to be cashed out not only in terms of its impact on health morbidity or life expectancy or pain, but also in terms of the effect it might have on prospects for achieving sufficiency with regard to reasoning, affiliation, security, respect and self determination. Thus, for example, the value of providing contraception includes not only its impact on a woman's prospects for surviving into older age, but also how control over whether and when to become a parent affects her prospects for living a life that is sufficiently self determining, with adequate social respect and sufficient capacity to care for those she loves. Also, in some cases, value can permissibly be cashed out not only in terms of the impact on the well being of those who are the direct recipients of a health good or service, but also on the well being of those who are caring for that person such as the loved ones of people with special needs.

The national entitlement may also include interventions that are higher cost or produce lower value, but only if in so doing there would be no negative impact on the availability of the good and services more essential to securing health sufficiency, or a sufficiency of the other core dimensions of well being, for all within its borders. Here, our theory mirrors, with a moral grounding rather than a strict efficiency grounding, the concern captured in allocational efficiency in economics to avoid expenditures and entitlements that "crowd out" other goods and services that maybe even more necessary for sufficiency of well-being or a decent human life.

Also, it is important to emphasize that sufficiency in health and the other dimensions is pegged to a decent life, and not an infinitely elastic notion of human flourishing. To do otherwise would result in another kind of unacceptable crowd out in which the pursuit of other valuable individual and social goods such as the arts and play and recreation would suffer.

PART THREE: NEW BIOTECHNOLOGIES AND THE EXAMPLE OF DEEP BRAIN STIMULATION

We turn now to the question with which this paper began: When, as a matter of justice, are societies obligated to include a new biotechnology in a national health care system or otherwise assure that all have reasonable access to it?

Consider again the example of DBS.

The possibility that DBS might become an effective treatment for memory disorders like Alzheimer's is extraordinarily important. Diseases and injuries that inter-

ferre with the cognitive processes that encode, store, and retrieve information threaten prospects for sufficiency in virtually every dimension of human well-being. As patients with progressive dementias lose the capacity to recall even the basic who's, what's, where's, and when's of their everyday lives, their suffering, and those of the people who love them, is enormous.

Much more work needs to be done to establish if DBS will prove effective in preventing or mitigating the ravages of Alzheimer's and other severe memory impairing disorders (Fell et al, 2012; Lee et al, 2013). However, assuming it does, and assuming that providing DBS to all relevant patients does not come at too high a price with regard to other sufficiency claims, then patients with Alzheimer's and the like would certainly be entitled to DBS under our sufficiency standard.

But also consider what we only hinted at in our initial discussion of the 2012 DBS epilepsy report. Some of the seven epilepsy patients in that experiment were NOT suffering from the memory impairment that sometimes co-travels with seizure disorders, yet *all* of them did better on the cognitive tasks when their brains were being stimulated. Assuming this line of research does succeed and the results are significant enough, it is a certainty that some people with normal cognitive function will seek DBS solely to get an intellectual edge or some other non-invasive neurotechnology such as transcranial magnetic stimulation.

We are now in the familiar and contested territory of the ethics of human enhancement, a complex set of issues that has generated a substantial literature (Buchanan, 2011; Savulescu & Bostrom, 2009; Harris, 2007; President's Council, 2003), much of it specific to cognitive enhancement including a limited literature on enhancement with invasive and non-invasive brain stimulation (Dresler et al, 2013; Kadosh et al, 2012; Hamilton et al, 2011). We engage here only one of those issues, the contentious relation between enhancement and justice, to which our account of sufficiency has at least a partial response.

Some argue that what is most morally troubling about enhancing biotechnologies is their potential to exacerbate or create egregious inequalities in human well-being. Some even hold that these technologies and social justice are on something like an inevitable collision course. Here we agree with Allen Buchanan's trenchant analysis that these criticisms are both overbroad, and misguided (Buchanan, 2011).

Even if it were possible to draw a sharp line between new biotechnologies that protect or restore health and new biotechnologies that enhance health or other dimensions of well-being, that difference does not itself necessarily pose fundamen-

tally different challenges for the kind of justice theory that we defend. In both cases, the same basic question is raised—as the technology diffuses, under what conditions, if ever, do individuals become entitled to it as a matter of justice?

Whether a biotechnology that enhances memory or any attribute critical to reasoning and cognitive performance should become a service to which individuals are entitled depends at least in part on the impact of access to that technology on considerations of relational equality. Put another way, a key question for us is whether, at some point in the diffusion of DBS in society, people without DBS would be so deprived in absolute terms, or systematically constrained relative to what is possible for others, that they cannot live decent lives as human beings or as members of a particular society.

The second aim of our theory, which we have only signaled here, focuses on the profound injustices that arise from conditions of systematic disadvantage. However, not all inequalities confer advantages and disadvantages, and even those that do confer advantages do not necessarily place others in a position of systematic disadvantage. What we are concerned about are the kinds of disadvantages that are so systematically oppressive that people cannot have decent lives.

Imagine a future in which so many people have DBS-enhanced memory such that, if you do not, your only options are to have the lowest, most menial jobs. *More importantly*, the negative impact of not being enhanced bleeds into many if not all of the other dimensions of well being. Without the enhancement, you cannot lead a self-determining life, you cannot understand the world around you well enough to be able to evaluate options and plan the way others can, you cannot assert your own interests in a meaningful way in the appropriate forums, and you cannot interact with others with a sense of your own comparable standing, but instead feel only shame or embarrassment.

If not having been treated with DBS renders you like the crossing sweeper Jo to the wealthy landowner John Jarndyce in Dicken's *Bleak House*, feeling unworthy of shelter in the other's person home even when deathly ill, then DBS for memory enhancement would fall within the entitlements of the sufficiency standard under our account.

Alternatively, imagine a future in which not having DBS is in its disadvantaging effects more analogous to not being able to attend an elite university. Although others with enhanced memory maybe advantaged in many arenas of life, much as the connections secured and arguably the education received by graduating from an

elite university is advantaging, your information processing skills are still adequate for participation in the economic and cultural life of society as a moral and political equal. You are still able to navigate the social world sufficiently to determine the broad pathways of your life and to advocate for your interests. You do not feel, nor do others perceive you, as less worthy of respect and dignified treatment because you have not been DBS-enhanced. In such a world, DBS for memory enhancement would fall above what sufficiency would require and would be ruled out as an entitlement.

Consider a further twist on the DBS example. Although we framed our original question in terms of national policy formulation, our theory and its implications extend as well to questions of global justice and access to new biotechnologies. Imagine that in affluent countries there is now universal access to DBS enhancement and that the global affluent also avail themselves of the intervention. The comparative disadvantages of not being enhanced, which we further imagine are as oppressive as was just described, now fall on the world's poor. Because of continued and escalating global interaction, these people are so systematically constrained relative to what is possible for others that they cannot live decent lives as human beings or as participants in the global order. DBS enhancement would now become required by our entitlement standard, transnationally as well as nationally.

In summary, the standards we set for identifying the goods and services that fall under the sufficiency threshold for health and the other dimensions of well-being are necessarily comparative. These standards are also necessarily dynamic, as dynamic as the creativity that is fueling the life and engineering sciences towards technologies that for some seem almost beyond imagining. Which technologies are required for sufficient well-being, for a decent life, will change over time and are in part context specific. The answer to that question, and to when on the diffusion of innovation curve the demands of justice are triggered, cannot be prescribed with algorithmic precision. A theory like ours can only provide broad guidance as to what should be taken into account in specific policy contexts. But there is no necessary reason to conclude that these answers, or the approach to these answers, have to be different when the technology is aimed at what some view as enhancing rather than protecting human biology, or even human nature.

REFERENCES

- Buchanan, A.E. 2010: *Human Rights, Legitimacy, and the Use of Force* (New York, NY: Oxford University Press).
- 2011: *Beyond Humanity?: The Ethics of Biomedical Enhancement* (New York, NY: Oxford University Press).
- Dresler, M., Sandberg, A., Ohla, K., Bublitz, C., Trenado, C., Mroczko-Wąsowicz, A., Kühn, S., and D. Repantis. 2013: 'Non-pharmacological cognitive enhancement'. *Neuropharmacology* 64: 529-543.
- Fell, J., Staresina, B.P., Do Lam, A.T.A. Widman, G., Helmstaedter, C., Elger, C.E., Axmacher N. 2013: 'Memory modulation by weak synchronous deep brain stimulation: A pilot study.' *Brain Stimulation* 6/3: 270-273.
- Hamilton, R., Messing, S., and A. Chatterjee. 2011: 'Rethinking the thinking cap: Ethics of neural enhancement using noninvasive brain stimulation'. *Neurology* 76(187-193).
- Harris, J. 2007: *Enhancing Evolution: The Ethical Case for Making Better People* (Princeton, NJ: Princeton University Press).
- Kadosh, R.C., Levy, N., O'Shea, J., Shea, N., and J. Savulescu. 2012: 'The neuroethics of non-invasive brain stimulation'. *Current Biology* 22:R108-R111.
- Lee, D.J., Gurkoff, G.G., Izadi, A., Berman, R.F., Ekstrom, A.D., Muizelaar, J.P., Lyeth, B.G., K. Shahlaie. 2013: 'Medial Septal Nucleus Theta Frequency Deep Brain Stimulation Improves Spatial Working Memory after Traumatic Brain Injury'. *Journal of Neurotrauma* 30: 131-139.
- Margalit, A. 1998: *The Decent Society* (Cambridge, MA: Harvard University Press).
- Mill, J.S. 1869: *The Subjection of Women* (London: Longmans, Green, Reader & Dyer).
- Nickel, J. 2007: *Making Sense of Human Rights* (Malden: Blackwell Publishing).
- Powers, M. and R. Faden. 2006: *Social Justice* (New York, NY: Oxford University Press).
- President's Council on Bioethics. 2003: *Beyond Therapy: Biotechnology and the Pursuit of Happiness* (Washington, DC: President's Council on Bioethics). Available at http://bioethics.georgetown.edu/pcbe/topics/beyond_index.html (accessed 26th March 2013).
- Savulescu, J. and N. Bostrom. (eds.) 2009: *Human Enhancement* (New York, NY: Oxford University Press).
- Suthana, N., Haneef, Z., Stern, J., Mukamel, R., Behnke, E., Knowlton, B. and I. Fried. 2012: 'Memory enhancement and deep-brain stimulation of the entorhinal area', *New England Journal of Medicine* 366/6: 502-10.

Situationism and Agency

ALFRED R. MELE

Department of Philosophy, Florida State University

JOSHUA SHEPHERD

Department of Philosophy, Florida State University

ABSTRACT

Research in psychology indicates that situations powerfully impact human behavior. Often, it seems, features of situations drive our behavior even when we remain unaware of these features or their influence. One response to this research is pessimism about human agency: human agents have little conscious control over their own behavior, and little insight into why they do what they do. In this paper we review classic and more recent studies indicating “the power of the situation,” and argue for a more optimistic response. In our view, though psychological research indicates situational influence, it also indicates that knowledge about the impact of situations on behavior can boost agents’ power to counteract harmful situational effects.



Any system of practical ethics makes some presuppositions about human agents. It may be assumed, for example, that many people are capable of making informed, conscious decisions about what to do in a wide range of situations and are capable of executing many such decisions. If people lack these capacities, the many practical injunctions that flow from ethical discourse come to seem misguided.

The assumption that people are capable of making informed, conscious decisions has been called into question by work in the human sciences. To take just one example, Benjamin Libet has argued, on the basis of his well-known neuroscientific findings, that, with one kind of exception, we never consciously decide what to do

(1985, 1999, 2004). The exception is for cases in which we become conscious of an urge or intention to do something. In some such cases, Libet claims, we are capable of consciously vetoing the urge or intention (1985, 1999, 2004).

One of us has argued that the neuroscientific case Libet presents for his thesis is far from persuasive (Mele 2009). Here we take up a related challenge to the common assumption mentioned in our opening paragraph and some related assumptions. The primary challenge we explore comes from the situationist (or “situationalist”) literature in psychology. This literature has received significant attention from philosophers in connection with moral virtue and moral character. Some see it as posing a serious challenge to the claim that there is such a thing as moral character (Doris 2002, Harman 1999), and others disagree (Kamtekar 2004, Sabini and Silver 2005, Sreenivasan 2002). We take a different approach here. Section One provides some background on situationism and briefly describes three well-known experiments on it. In section two, we contrast two perspectives on this literature, one pessimistic and another optimistic, and we side with the optimists. In our view, knowledge about the influences of situations on behavior can boost agents’ power to counteract harmful situational effects. In section three, we review recent work indicating that social cues also strongly influence behavior in a way that is mediated by unconscious processes. Again, we offer an optimistic interpretation of the import of this work for our capacity to control our behavior, and we provide some empirical support for our interpretation. In section four, we discuss some additional evidence that supports our optimism about agential control. The upshot is that although unconscious processing certainly influences behavior, there is reason to think that education about the influence of pertinent stimuli can play an ameliorating role in cases of undesirable influence. Section Five wraps things up.

SECTION ONE: SOME CLASSIC SITUATIONIST FINDINGS

Matthew Lieberman writes:

If a social psychologist was going to be marooned on a deserted island and could only take one principle of social psychology with him it would undoubtedly be “the power of the situation.” All of the most classic studies in the early days of social psychology demonstrated that situations can exert a powerful force over the actions of individuals. . . . If the power of the situation is the first principle of social psychology,

a second principle is that people are largely unaware of the influence of situations on behavior, whether it is their own or someone else's behavior. (2005, p. 746)

In the present section we describe three of these classic studies.

In a study by John Darley and Bibb Latané (1968), participants were led to believe that they would be talking about personal problems associated with being a college student. Each was in a room alone, thinking that he or she was talking to other participants over a microphone. Sometimes participants were led to believe that there was only one other participant (group A), sometimes that there were two others (group B), and sometimes that there were five others (group C). In fact, the voices the participants heard were recordings. Participants were told that while one person was talking, the microphone arrangement would not let anyone else talk. At some point, the participant would hear a person—the “victim”—say that he felt like he was about to have a seizure. The victim asks for help, rambles a bit, says he is afraid he might die, and so on. His voice is abruptly cut off after he talks for 125 seconds, just after he makes choking sounds. The percentage figures for participants who left the cubicle to help before the voice was cut off are as follows: group A 85%, group B 62%, group C 31%. Also, all the participants in group A eventually reported the emergency, whereas only 62% of the participants in group C did this.

Clearly, participants' beliefs about how other many other people could hear the voice—none, one, or four—had an effect on their behavior. Even so, there being one or four other people around to help the victim seems not to be a reason not to help.

Philip Zimbardo's Stanford prison experiment (Haney et al. 1973, Zimbardo n.d., Zimbardo et al. 1973) began with newspaper ads for male college students willing to take part in an experiment on prison life. The volunteers selected as prisoners were arrested at their residences, handcuffed, searched, and driven in a police car to a Palo Alto police station. From there, after being fingerprinted and placed in a detention cell, they were driven to a mock prison built in the basement of the Stanford psychology building. When they arrived, they were stripped and sprayed with deodorant. Then, after being given a prison uniform and photographed, they were locked in cells. There were three small cells—six by nine feet—for the ten prisoners and a very small solitary confinement cell. There were also rooms for volunteers selected as guards. Much of the activity was videoed by hidden cameras. Concealed microphones picked up conversations.

The plan was to run the experiment for two weeks. The prisoners were there twenty-four hours per day. The guards worked eight-hour shifts and then went home. Prisoners had three simple, bland meals a day and the same number of supervised toilet visits. They were also lined up three times each day to be counted and were always referred to by a number worn on their uniform—never by their name. They had two hours of free time each day to write letters or read—unless that privilege was taken away. And they had chores to do—cleaning toilets and the like. It is interesting that during their free time, 90% of what the prisoners talked about had to do with their prison life.

Zimbardo and coauthors report that “five prisoners had to be released because of extreme emotional depression, crying, rage and acute anxiety” (Haney et al. 1973, p. 81). Although the experiment was supposed to last two weeks, Zimbardo ended it after just six days. One prisoner had to be released after thirty-six hours owing to “extreme depression, disorganized thinking, uncontrollable crying and fits of rage” (Zimbardo et al. 1973). Another developed a psychosomatic rash.

Several of the guards became bullies, and those who did not participate in the bullying allowed it to continue. The harassment increased each day. Counting of prisoners, which originally took ten minutes, sometimes went on for hours. During these counts, prisoners were encouraged to belittle each other. Over time, the prisoners’ attitude toward one another reflected the guards’ attitude toward them. Insults and threats escalated, and so did commands to do pointless or demeaning tasks. Guards sometimes made prisoners clean toilets with their bare hands.

Pointless tasks included moving boxes back and forth from one closet to another and picking thorns out of blankets after guards had dragged the blankets through bushes. Sometimes prisoners would be made to do push ups while guards stepped on them. Guards would wake prisoners up in the middle of the night to count them. Sometimes they would deny them their scheduled leisure time just for the fun of it, or lock them in a solitary confinement cell for no good reason—a seven foot tall broom closet two feet wide and two feet deep. After the 10:00 p.m. lockup, prisoners often had to use buckets in their cells as toilets. On the second day of the experiment, prisoners staged a protest. The guards used a fire extinguisher to spray them, stripped them, and put the leaders in solitary confinement.

The guards created a privilege cell to sow dissension among the prisoners. The good prisoners would use the cell and get better treatment, including better food. After a while, to confuse the prisoners, the ones who seemed worse got the

privileges. Some of the guards became sadistic; and, of course, Zimbardo was as interested in the effects on the guards as the effects on the prisoners.

Bad effects of the situation showed up both in prisoners and in guards. The guards fell into three types. Some were tough but fair, some were good guys who did small favors for prisoners, and about a third were hostile and abusive. None of the testing the experimenters did in advance predicted which of the students would become power-loving guards. Some of the guards were disappointed that the experiment ended early; they enjoyed their power.

One of the prisoners felt sick and wanted to be released. He cried hysterically while talking with Zimbardo, in his role as prison superintendent, and a priest. After Zimbardo left the room to get the prisoner some food, the other prisoners began to chant that this one was a bad prisoner. When Zimbardo realized that the prisoner could hear this, he ran back into the room. He writes:

I suggested we leave, but he refused. Through his tears, he said he could not leave because the others had labeled him a bad prisoner. Even though he was feeling sick, he wanted to go back and prove he was not a bad prisoner. At that point I said, 'Listen, you are not #819. You are [his name], and my name is Dr. Zimbardo. I am a psychologist, not a prison superintendent, and this is not a real prison. This is just an experiment, and those are students, not prisoners, just like you. Let's go.' He stopped crying suddenly, looked up at me like a small child awakened from a nightmare, and replied, 'Okay, let's go'. (Zimbardo n.d.)

This episode makes especially salient how deeply participants were drawn into their roles.

We turn to Stanley Milgram's famous studies of obedience, beginning with the study first reported in Milgram 1963. Participants were forty men between the ages of twenty and fifty and from many different walks of life. The cover story was that the experimenter was conducting an experiment on how punishment is related to memory.

The participant meets the experimenter and a confederate of his. The participant is told that he and the other man will draw slips of paper from a hat to see which of them will be the "teacher" and which the "learner." In fact, the participant is always the teacher. He hears the cover story and sees where the learner will sit during the experiment—in a chair in which the learner will supposedly receive an electric

shock from the teacher each time he gives an incorrect answer. The teacher watches the learner being strapped into the electric chair and is told the straps will prevent him from moving too much when he is being shocked. The teacher then moves to another room where he can no longer see the learner. Milgram reports that, with a few exceptions, participants believed the setup was real.

Participants are shown an array of thirty levers, each associated with different degrees of shock. The lowest shock is for the first incorrect answer, the second lowest is for the second wrong answer, and so on. Sets of levers—mainly sets of four—are labeled. About half way through, the label is “intense shock,” followed by “extreme intensity shock,” “danger: severe shock,” and finally “XXX.”

The learner answers by pressing a button. At one point during the experiment—after he has received his twentieth shock—the learner pounds on the wall, and from then on he does not answer any more questions. The twentieth shock was delivered by the fourth lever in the “intense shock” level. The shock levels were also labeled with voltage numbers. This one was 300 volts. Before shocking the learner, the teacher had to report the voltage of the shock he was about to administer: 15 at the beginning all the way up to 450 at the end. At the beginning of the experiment, the scientist told the teacher that “Although the shocks can be extremely painful, they cause no permanent tissue damage.” When participants raised the issue of stopping the experiment, they were given stock replies ranging from “Please continue” to “You have no other choice, you must go on.” The scientist started with a simple request to continue and eventually moved up to the “no choice” response if the participant persisted in talking about stopping.

Twenty-six of the forty participants continued shocking all the way to the end. (Teachers were told that no answer counted as a wrong answer.) No participant stopped shocking before the twentieth shock. Five stopped right after that one. Four stopped after the next one: it was the first shock in the series labeled “extreme intensity shock” and the first shock in response to a non-answer. The other four dropped out a bit later.

Milgram reports that the participants displayed enormous tension, fits of nervous laughter, twitching, stuttering, sweating, and the like. And when they talked about stopping, a calm reply by the experimenter often worked: “The experiment requires that you continue,” “It is absolutely essential that you continue,” or the like. If a participant refused to continue after being told he had no choice, the experiment

was terminated and the participant was debriefed. This “no choice” response was the last in a series of four stock responses by the experimenter.

Milgram conducted many versions of this experiment. A brief description of three additional versions will prove useful. In Voice-Feedback (Milgram 1974, experiment 2), the teacher could now hear the learner speak. The learner grunts in response to the 75 volt shock and the slightly later ones. At 120 volts—labeled moderate—he shouts and says the shocks are becoming painful. He groans after the next shock, and refuses to continue in response to the one after that—the tenth shock. This goes on with increasing intensity for several more shocks. At 180 volts, the learner screams that he cannot stand the pain. By 270 volts he is screaming in agony. At 300 volts—the twentieth shock—he desperately shouts that he will not provide any more answers. And he repeats this after the next shock—after emitting a violent scream. After all subsequent shocks, he shrieks in agony. Twenty-five of the forty participants shocked all the way to the end.

In two other versions of the experiment, the teacher was brought much closer to the learner, but everything else was very similar—the groaning, screaming, and so on. There were forty participants in each. In one version (Proximity: Milgram 1974, experiment 3) the teacher was just a foot and a half from the learner and could see him clearly. In the other (Touch-Proximity: Milgram 1974, experiment 4), the learner could remove his hand from a shock plate in order to avoid being shocked, and the teacher would have to force the learner’s hand onto the plate in order to shock him. There were forty participants in each. In Proximity, sixteen participants continued to the end. In Proximity-Touch, twelve did.

SECTION TWO: PERSPECTIVES ON THE CLASSIC STUDIES

The findings we described (along with findings of many related studies) certainly are interesting. What should we make of them? According to a pessimistic view, they suggest that people have very little control over their behavior—that human behavior is largely driven by the situations in which people find themselves and the effects these situations have on automatic behavior-producing processes.

We are not so pessimistic. A few days after the tragic events of September 11, 2001, a friend said “That will never happen again.” He explained that, in his view, people would learn from what happened, and, henceforth, a plane full of passengers would not go down without a fight. They would resist, and they would overpower

their foes. That was an uplifting thought (a thought inspired in part by news reports that passengers and crew on United Airlines flight 93 had heard about the earlier crashes and attempted to regain control of the plane).

The role of “passenger” on a commercial flight is pretty well defined. Passengers are to sit down, fasten their seat belts, keep them fastened until they are told they are permitted to get up, refrain from being disruptive in any way, and, in general, obey the airline employees. For the most part, if there is a disturbance, passengers expect the flight crew to deal with it. The passengers’ situation involves ingredients of the three studies we described. The prisoners and guards occupy a role in Zimbardo’s studies; so do passengers. Obedience to pertinent authority figures is something typical passengers share with typical participants in Milgram’s studies. And when there is a disturbance on a plane, nonintervention by passengers is not surprising, especially given that such disturbances are matters the airline employees are expected to handle. In the bystander study that we described, participants had no reason to believe that an authority figure (the experimenter) was aware of the apparent emergency. So nonintervention by airplane passengers would seem to be more predictable, other things being equal.

Now, if behavior is driven by situations in such a way that new, consciously processed information is out of the behavior-producing loop, then our friend was way too optimistic. But we are inclined to agree with him. If we had had the horrible misfortune to be on one of the airliners that hit the World Trade Center years ago, we probably would have refrained from intervening and hoped that the airline employees would handle things. In light of what we learned, we predict that our reactions would be different now. Our expectation is that if a passenger or two attempted to intervene, others would join in.

This last remark is a window on our optimism. Behavioral education starts at an early age. Parents try to teach their toddlers to control potentially harmful impulses, and they enjoy a considerable measure of success. Parents also teach respect for parental authority; and they engage in moral education, which also involves instruction in self-control. Of course, parents can only teach what they are familiar with. And a lot more is known now about factors that influence human behavior than was known fifty years ago. Our view is that this knowledge should be put to good use, and not only in child rearing.

One often sees articles for public consumption claiming that neuroscientists have shown that free will is an illusion. One of us has made various attempts to

debunk such claims (see, e.g., Mele 2009 and Mele n.d.), but our point now is that lots of people find striking “news” about human behavior interesting. The classic studies that we described are not news now, of course; but they continue to be cited in new studies on situationism or automaticity. One way to spin news about these stories is pessimistic: for example, being in a group that witnesses an emergency has an enormous effect on your behavior, and there is nothing you can do about it. Another is not: and now that you know about the bystander effect, what will you do should you find yourself in a group that witnesses an emergency?

There are plenty of self-help books on self-control. People learn techniques for resisting or avoiding temptation with a view to making their lives go better. People who read such books know what they want to avoid—binge eating, gambling, binge drinking, or whatever—and they try to learn how to avoid it. When a cause of harmful behavior flies under everyone’s radar, not much can be done about it. But once a cause of harmful action or inaction is brought to light, prospects for amelioration may become brighter.

A public that is educated about the bystander effect is less likely to display it. The same is true of undue or excessive obedience to authority. In the latter sphere, matters are delicate. Obedience to authority is important for civil society. Because it is useful, it is instilled by parents, teachers, and so on; and it tends to become habitual in many people. But we also know the evils to which it can lead. Milgram’s work was motivated partly by a desire to understand how ordinary German citizens who became rank and file military personnel ended up committing atrocities. Obedience to authority is an important part of his answer. It would seem that the socialization of obedience to authority should include education about proper limits to obedience. Milgram writes: “In growing up, the normal individual has learned to check the expression of aggressive impulses. But the culture has failed, almost entirely, in inculcating internal controls on actions that have their origin in authority. For this reason, the latter constitutes a far greater danger to human survival” (1974, p. 147). Education can lessen this danger.

What about Zimbardo’s findings? They have obvious implications for the training of prison guards, and the implications clearly extend to people whose jobs give them considerable power over others—police, for example. But the import of his findings extends much further. There are situations in which continuing to play whatever role we are playing at the time—passenger, army private, student—will

handicap us. The knowledge that that is so can make it easier for us to shed our roles when the time is right to do that.

SECTION THREE: SOME RECENT FINDINGS

Recent findings about unconscious influences on behavior are interesting too. In this section, we review evidence of these influences. While a superficial reading of this evidence suggests a pessimistic view about the human capacity for self-control, we offer grounds for optimism.

Melissa Bateson and colleagues (Bateson et al. 2006) conducted an experiment in the office of the Psychology Department at the University of Newcastle. The office keeps coffee, tea, and milk on hand. Department members pay for the drinks by voluntarily depositing money in an “honesty box.” Bateson and colleagues tweaked this system in an interesting way. On a cupboard door located above the honesty box and the drink-making supplies, they posted an instruction sheet with the following suggestions: 30 pence for tea, 50 pence for coffee, 10 pence for milk. In addition to these suggestions, the sheet included an image: either a pair of eyes looking at the observer or flowers. The experiment ran for ten weeks. Each week the experimenters switched the image and recorded the amount of money given that week.

Contributions to the honesty box reliably tracked the change in images. Each time the experimenters replaced the flowers with watching eyes, contributions rose. And each time they replaced the watching eyes with flowers, contributions dropped. On average, department members contributed 2.76 times more money when the eyes were watching (Bateson et al. 2006, p. 412).

Why would the image of eyes have such a powerful effect on behavior? Bateson and colleagues speculate that the eyes “induce a perception in participants of being watched” (2006, p. 413). Importantly, based on information that the human perceptual system is highly sensitive to social stimuli such as eyes, they speculate that this perception is largely non-conscious. The idea is that the presence of the watching eyes activates non-conscious “reputational concerns” that motivate increased contributions (p. 413).

This interpretation receives some support from a study conducted by Mary Rigdon and colleagues. Rigdon et al. 2009 had participants play a version of the “dictator game” in which one participant (the Dictator) was given \$10 and told to indicate on a decision sheet how much—in \$1 increments—he or she wished to give to

another participant (the Recipient). Because it is well-known that social cues about Recipients (e.g., being told the Recipient's surname) influence how much Dictators give, in this study Dictators and Recipients were kept anonymous. The only social cue the Dictators received came in the form of three dots, located in the center of the decision sheet, just above the place where they were to indicate the amount they would give. One group saw three dots arranged to resemble watching eyes and a nose: two dots on top, one the on bottom. A second group saw the three dots arranged in a neutral configuration: one dot on top, two dots on the bottom. This is obviously a very minimal social cue. But dots arranged in the watching eyes configuration are known to activate the part of the brain responsible for face recognition, the fusiform face area (Tong et al. 2000); and Rigdon et al. hypothesized that even this minimal cue would influence the amount given by Dictators.

They were right, but with a twist. Male participants gave significantly more when they saw the watching eyes configuration of dots. On average, male Dictators in the watching eyes condition gave \$3.00; male Dictators in the neutral dots condition gave \$1.41. Furthermore, male Dictators in the watching eyes condition gave \$1 or more 79% of the time, compared with only 37% in the neutral dots condition (Rigdon et al. 2009, p. 362). It seems that for many males, the three dots totally changed their giving behavior.

Female participants' giving patterns were not influenced by the dots at all. Rigdon et al. explain the difference by pointing to other studies using the dictator game that indicate that, on the whole, female Dictators give much more than male Dictators. Since female Dictators "seem to already view the choice problem . . . as a social allocation task" the watching eyes should not stimulate pro-social behavior (2009, p. 363). The opposite is true for male Dictators, however. They tend to use their anonymity to their own economic advantage. Thus, for male Dictators, "Processing the stimulus ultimately activates the fusiform face area of the brain, making the environment seem—at a pre-conscious level, perhaps accessible to the decision-making process but not to introspection—less anonymous and hence less socially distant" (p. 363).

These kinds of studies certainly seem to favor a pessimistic view regarding the human capacity for self-control. Having your fusiform face area stimulated by three dots that vaguely resemble eyes and a nose might activate non-conscious processes that cause you to give money when otherwise you would not have—especially if you were already inclined to be stingy. However, as we noted in section two, it seems

that knowledge about these kinds of non-conscious influences has the potential to enhance our capacity to counteract them. Someone who knows that the presence of eyes, or even the feeling of being watched, tends to cause certain patterns of behavior, might be able to counteract the non-conscious influence. Imagine receiving a flyer in the mail asking for your donation to some politician's campaign, and imagine noticing the image of a face or of eyes in the center of the page. Now that you know that perceiving eyes influences giving behavior, it is possible that different processes will be activated in you. Perhaps you will think more deliberately about how much money you have to donate this month. Perhaps, recognizing an attempt at manipulation, you will toss the flyer into the trash. That is up to you. Our point is that knowledge about non-conscious processes that influence behavior has the potential to mitigate the influence those processes have by activating more explicit processes.

Is there hard evidence that supports our optimism about knowledge? An interesting line of research concerns the behavioral influence of implicit attitudes—attitudes agents possess, but of which they are rarely aware. The most popular measure of such attitudes is the implicit association test (IAT). On a typical IAT—for example, one measuring implicit attitudes towards black and white people—a participant sits in front of a computer screen and is asked to categorize stimuli by pressing one of two keys. There are four types of stimuli—black people's faces, white people's faces, negative words, and positive words—and two response keys. For each response key, researchers pair a black face with either a positive or a negative word and a white face with either a positive or a negative word. This yields two types of response situation: "compatible" (i.e., black face/negative word or white face/positive word) and "incompatible" (i.e., black face/positive word or white face/negative word). Examples of positive words are "joy," "laughter," "love," and "peace"; negative words include "evil," "failure," "nasty," and "terrible." Participants see a series of stimuli and react appropriately: for example, black person's face (press left key), positive word (press left key), negative word (press right key), white person's face (press right key). Researchers then measure how long it takes participants to press the relevant key in different response situations. If they find, as they often do, that it takes longer to categorize a black person's face when the response key maps a black face with a positive word, they conclude that an implicitly negative attitude towards black people exists. As Fiedler and Bluemke explain, "Whoever participated in an IAT, swearing not to be prejudiced at all against Blacks, will have found it nevertheless much easier to use the same re-

sponse for White and positive and for Black and negative than vice versa” (2005, p. 307).

It is commonly assumed that the influence of implicit attitudes on behavior is important and occurs largely by way of non-conscious processing. To take one example, Neil Levy remarks that “implicit attitudes probably explain some incidents involving lethal force. Priming with black faces raises the likelihood that agents will identify ambiguous stimuli or non-gun tools as guns . . . This fact may partially explain why police are more likely to use deadly force when confronted with black suspects” (2012, p. 9).

The assumption that implicit attitudes influence behavior in ways beyond an agent’s conscious control is reflected in assumptions about the IAT. It is assumed that at very short time scales of one second or less, participants cannot consciously control their reactions to stimuli. As Fiedler and Bluemke note, “It is this apparent lack of control or impossibility to counteract the IAT effect that has nourished the claim that an unobtrusive instrument [for measuring implicit attitudes] has been found, which does not lend itself to controlled responding” (2005, p. 307).

The assumption at issue about conscious control is false. Fiedler and Bluemke (2005) gave German participants an IAT that measured negative attitudes against Turks and then asked them to take it again and to try “to avoid a result that would indicate a negative implicit attitude against Turks” (p. 308). Note that even if participants actually have implicitly negative attitudes against Turks, there are, in principle, two ways they might avoid such a result. First, they might slow down their responses in compatible response situations (e.g., Turkish face/negative word). Second, they might speed up their responses in incompatible response situations (e.g., Turkish face/positive word). The slowing down response is one way to exercise control over the influence of one’s implicit attitude: one “beats the test” by pretending that it takes one just as long to categorize stimuli in the compatible as in the incompatible condition. The speeding up strategy is another way to exercise control over the influence of one’s attitude: by making a successful conscious effort, one nullifies the influence of one’s implicit attitude on one’s response time in incompatible response situations.

Fiedler and Bluemke’s participants seemed to utilize both strategies. On their second time through the IAT, they slowed down responses in compatible response situations and sped up responses in incompatible situations (2005, p. 310, Table 2). Fiedler and Bluemke did not predict the latter result, and they were surprised that participants were able to speed up responses (p. 315). They suggested that the speed

up might have been due to practice alone (p. 315). An alternative possibility, not considered by Fiedler and Bluemke, is that participants' familiarity with the test, coupled with an intention to speed up responses, led to the speed up. And there is evidence that bears on it, as we will explain.

Xiaoqing Hu and colleagues (2012) had participants take an IAT and then take it again. On the second trial, they separated participants into four groups. Group 1 simply repeated the IAT to test for the influence of task repetition. Group 2 repeated the incompatible response block of the IAT three times to test for the influence of practice. Group 3 was explicitly instructed to speed up their responses in incompatible response situations. Group 4 was told the same thing as group 3, and they were also given more time to practice; they repeated the incompatible response block three times, just like group 2.

If a conscious intention to speed up responses is to be effective, one would expect group 3 to respond faster than group 1 in the incompatible response conditions. One would also expect group 4 to respond faster than group 2 in the incompatible response conditions. This is what happened (Hu et al. 2012, p. 3, Table 1). Group 3 improved response time by 168 ms (from 902 ms to 734 ms), while group 1 improved response time only by 45 ms (from 950 ms to 905 ms). Compared with group 2, group 4 significantly improved response time as well. Practice certainly seemed to help: group 2 improved response time by 80 ms (from 922 ms to 842 ms). But group 4 improved response time by 215 ms (from 858 ms to 643 ms).

That both a conscious intention and training in speeding up responses had large effects on behavior constitutes important evidence in favor of our optimism. Participants were, in effect, asked to control the influence of implicit attitudes on behavior at a very rapid time scale—less than a second. Participants informed about the influence of implicit attitudes on behavior were able to successfully control the influence of these implicit attitudes. This directly counters the common assumption that implicit attitudes influence behavior in ways not susceptible to conscious control. Knowledge about effects on agents that normally fly under the radar of agents' consciousness can give people the power to weaken those effects. The fact that relevant knowledge can do this at such rapid time scales is striking, and it speaks against a pessimistic perspective on agential control.

SECTION FOUR: IMPLEMENTATION INTENTIONS
AND THE ZOMBIE HYPOTHESIS

John Kihlstrom reports that some social psychologists “embrace and promote the idea that automatic processes dominate human experience, thought, and action to the virtual exclusion of everything else” (2008, p. 168). If he is exaggerating, he is not exaggerating much: for example, Daniel Wegner contends that conscious intentions are never among the causes of corresponding actions (2002, 2004, 2008). However, there is additional good evidence that conscious intentions do important work. If this were a book, we might try to catalogue such evidence. Instead we will concentrate on a particular body of work with clear connections to self-control. We want to continue countering the impression that, as science correspondent Sandra Blakeslee put it in a New York Times article, “in navigating the world and deciding what is rewarding, humans are closer to zombies than sentient beings much of the time” (as quoted in Kihlstrom 2008, p. 163).

Some of Milgram’s descriptions of the excessively obedient behavior he observed are similar to descriptions of akratic action, sometimes defined as un-compelled, intentional action contrary to the agent’s conscious better judgment (see Mele 2012, p. 3). For example, he asserts that “Some subjects were totally convinced of the wrongness of what they were doing” (1974, p. 10) and that “many subjects [who continue shocking] make the intellectual decision that they should not give any more shocks” (p. 148). It is possible that Milgram did not have akratic action in mind. He might have thought that the participants at issue were compelled to act as they did. Milgram completes the first sentence quoted in this paragraph with the words “but *could not* bring themselves to make an open break with authority,” and the second sentence ends with the words “they are frequently *unable* to transform this conviction into action.” (The emphasis is ours in both cases.) But he might have meant that the participants found it very difficult to do what they believed was right and failed to do the right thing.

The flip side of akratic action is enkratic action, action exhibiting self-control in the face of pressure to act contrary to one’s better judgment. A large body of work on “implementation intentions” provides encouragement concerning our prospects for self-control (for reviews, see Gollwitzer 1999 and Gollwitzer and Sheeran 2006) while also countering the idea that conscious intentions have virtually no effect on intentional action.

Implementation intentions, as Peter Gollwitzer describes them, “are subordinate to goal intentions and specify the when, where, and how of responses leading to goal attainment” (1999, p. 494). They “serve the purpose of promoting the attainment of the goal specified in the goal intention.” In forming an implementation intention, “the person commits himself or herself to respond to a certain situation in a certain manner.”

In one study of participants “who had reported strong goal intentions to perform a BSE [breast self-examination] during the next month, 100% did so if they had been induced to form additional implementation intentions” (Gollwitzer 1999, p. 496). In a control group of people who also reported strong goal intentions to do this but were not asked to form implementation intentions, only 53% performed a BSE. Participants in the former group were asked to state in writing “where and when” they would perform a BSE during the next month. These statements expressed implementation intentions.

The featured future task in another study was “vigorous exercise for 20 minutes during the next week” (Gollwitzer 1999, p. 496). “A motivational intervention that focused on increasing self-efficacy to exercise, the perceived severity of and vulnerability to coronary heart disease, and the expectation that exercise will reduce the risk of coronary heart disease raised compliance from 29% to only 39%.” When this intervention was paired with the instruction to form relevant implementation intentions, “the compliance rate rose to 91%.”

In a third study reviewed in Gollwitzer 1999, drug addicts who showed symptoms of withdrawal were divided into two groups. “One group was asked in the morning to form the goal intention to write a short curriculum vitae before 5:00 p.m. and to add implementation intentions that specified when and where they would write it” (p. 496). The other participants were asked “to form the same goal intention but with irrelevant implementation intentions (i.e., they were asked to specify when they would eat lunch and where they would sit).” Once again, the results are striking: although none of the people in the second group completed the task, 80% of the people in the first group completed it.

Many studies of this kind are reviewed in Gollwitzer 1999, and Gollwitzer and Paschal Sheeran report that “findings from 94 independent tests showed that implementation intentions had a positive effect of medium-to-large magnitude . . . on goal attainment” (2006, p. 69). These results provide evidence that the presence of relevant distal implementation intentions significantly increases the probability that agents

will execute associated distal “goal intentions” in a broad range of circumstances. In the experimental studies that Gollwitzer reviews, participants are explicitly asked to form relevant implementation intentions, and the intentions at issue are consciously expressed (1999, p. 501). (It should not be assumed, incidentally, that all members of all of the control groups lack conscious implementation intentions. Indeed, for all anyone knows, many members of the control groups who executed their goal intentions consciously made relevant distal implementation decisions.)

Research on implementation intentions certainly suggests that one useful technique for mastering anticipated motivation not to do what one judges it best to do later—for example, exercise next week or finish writing a C.V. by the end of the day—is simply to decide, shortly after making the judgment, on a very specific plan for so doing. Of course, what works against relatively modest motivational opposition might not work when the opposition is considerably stronger, as it may often be in the case of addicts’ desires for their preferred drugs (see Webb et al. 2009).

We are not suggesting that implementation intentions provide a solution to the problems encountered by participants in the classic studies we have discussed. Our purpose in this section has been to offer some grounds for not being overly impressed by the zombie hypothesis about human beings and some support for our optimism about human prospects for self-control (for additional support, see Mele 2012, ch. 5). The key to dealing with the bystander effect, the power of roles, and excessive obedience, we have suggested, is education. Sometimes, knowledge is power.

SECTION FIVE: OUR STRATEGY

A brief discussion of our strategy in this article is in order. We began by noting that the assumption that people are capable of making and acting on informed, conscious decisions has been challenged by work in the human sciences. The primary challenge that we selected for discussion comes from the situationist literature in psychology. In sections 1 and 2 we reviewed some classic situationist findings and sketched a case for an optimistic perspective on them. Readers will have noticed that our optimism has a cautious tone: for example, we claimed that “once a cause of harmful action or inaction is brought to light, prospects for amelioration may become brighter.” We suggested that a public that is educated about the bystander effect is “less likely to display it,” and we made comparable suggestions about undue or excessive obedience to authority and the effects of agents’ roles.

In section three, we turned to recent findings. We had two aims there. One was to move the discussion into the sphere of current scientific research. The other and more important aim was to call attention to some hard evidence that knowledge about the effects of unconscious processes on behavior can help people to counteract those effects. The work on implicit attitudes that we discussed was useful in this connection, even though it is not situated in the situationist literature. One looks for evidence where one thinks one has a good chance of finding it.

An important part of the assumption that has been our topic in this article is that conscious intentions and decisions can have an effect on behavior, and, more specifically, that they (or their physical correlates) can issue in corresponding actions. The most vigorous opponent of this assumption is Daniel Wegner (2002, 2004, 2008). Our primary aim in section four was to present hard evidence that the opposition is mistaken—evidence from research on implementation intentions. This research is not situated in the situationist tradition; but, again, it makes sense to look for evidence where one believes one is likely to find it.

At this point, the discussion could have taken a metaphysical turn. Is it conscious implementation intentions qua conscious intentions that do the work or is the work instead done by their physical correlates? We opted against discussing that issue here, and we direct readers interested in the topic to Mele 2009, pp. 146-48.

An empirical issue that we did not pursue merits at least a brief mention. There is evidence that pertinent implementation intentions reduce automatic stereotyping (Stewart and Payne 2008). This finding links our discussion of implementation intentions in section four to our discussion of implicit attitudes in section three. (See Fine 2006 for a review of evidence that “automatic social processes can come to be importantly constrained by prior controlled cognitive processes” (p. 85), including prior processes involving implementation intentions (pp. 92-93).) But, again, our primary aim in section four was to counter the idea that conscious intentions play no role in the production of corresponding actions and that automaticity rules here.

We chose not to critique various studies of unconscious influences on behavior. But we would be remiss if we did not note the existence of relevant critiques. For example, in a review of research on unconscious influences on decision making, Ben Newell and David Shanks argue that in various alleged demonstrations of such influence, “inadequate procedures for assessing awareness, failures to consider artifactual explanations of ‘landmark’ results, and a tendency to uncritically accept conclusions that fit with our intuitions have all contributed to unconscious influences being as-

cribed inflated and erroneous explanatory power in theories of decision making” (n.d., p.1).

As we see it, good scientific research on the effects of unconscious processes on behavior should be encouraged, as should good critiques of that work. We value knowledge of the springs of human behavior for its own sake, but such knowledge has instrumental value as well, including the value that the cautiously optimistic perspective we have developed highlights. Sometimes, we said, knowledge is power. Here is another way to put it: sometimes, forewarned is forearmed. It is knowledge about actual broad effects that we have in mind—for example, the bystander effect. We know of no direct evidence that informing people about the bystander effect can influence their behavior in bystander situations. But we do believe that people should be informed about the effect, and we hope that well-presented information will have a positive influence on behavior. In fact, this is part of our reason for describing the bystander study in section one, and we had similar motivation for our presentation of the other classic studies described there. Our understanding is that the readership of this journal will include people who have only a vague familiarity with situationism and classic situationist studies, and the *Journal of Practical Ethics* is a good place for theoretical papers that also pursue some practical aims with ethical significance.

Someday, we may try to catalogue harmful or counterproductive unconscious processes that can be consciously counteracted. We have discussed implicit attitudes in this connection and we have mentioned that implementation intentions have been found to reduce automatic racial stereotyping. Here are two more examples. First, there is evidence that unconscious gender bias in hiring decisions can be counteracted by consciously settling on hiring criteria before the candidates’ gender is disclosed (Uhlmann and Cohen 2005). Second, the confirmation bias—the tendency to search (in memory and the world) more often for confirming than for disconfirming instances of a hypothesis one is testing and to recognize confirming instances more readily—can be counteracted by consciously taking the perspective of someone whose job it is to find violations of a rule (Gigerenzer and Hug 1992). We find research on unconscious processes useful both for what it tells us about how human beings function and for what it might tell us about how human beings can function better. Of course, we are by no means suggesting that most unconscious processes are counterproductive. Some are; many are very useful.

CONCLUSION

The classic situationist studies that we discussed are disconcerting. One response is pessimism about human agency: some may conclude that intentional human action is driven primarily by forces that fly under the radar of consciousness and that we have little insight, as agents, into why we do what we do. Not only have we discussed some evidence to the contrary, but we also have provided grounds for an optimistic view according to which knowledge about situational influences can improve human agents' prospects for dealing rationally with them.

Acknowledgments: We are grateful to two anonymous referees for useful comments on an earlier version of this article. This article was made possible through the support of a grant to Mele from the John Templeton Foundation. The opinions expressed in this article are our own and do not necessarily reflect the views of the John Templeton Foundation.

REFERENCES

- Bateson, M., D. Nettle, and G. Roberts. 2006. Cues of Being Watched Enhance Cooperation in a Real-World Setting. *Biology Letters*, 2: 412-414.
- Darley, J., and B. Latané. 1968. Bystander Intervention in Emergencies: Diffusion of Responsibility. *Journal of Personality and Social Psychology*, 8: 377-383.
- Doris, J. 2002. *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.
- Fiedler, K. and M. Bluemke. 2005. Faking the IAT: Aided and Unaided Response Control on the Implicit Association Tests. *Basic and Applied Social Psychology*, 27: 307-316.
- Fine, C. 2006. Is the Emotional Dog Wagging its Rational Tail, or Chasing it?. *Philosophical Explorations*, 9: 83-98.
- Gigerenzer, G. and K. Hug. 1992. Domain-Specific Reasoning: Social Contracts, Cheating, and Perspective Change. *Cognition*, 43: 127-171.
- Gollwitzer, P. 1999. Implementation Intentions. *American Psychologist*, 54: 493-503.
- Gollwitzer, P. and P. Sheeran. 2006. Implementation Intentions and Goal Achievement: A Meta-Analysis of Effects and Processes. *Advances in Experimental Social Psychology*, 38: 69-119.
- Haney, C., W. Banks, and P. Zimbardo. 1973. Interpersonal Dynamics of a Simulated Prison. *International Journal of Criminology and Penology*, 1: 69-97.

Harman, G. 1999. Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error. *Proceedings of the Aristotelian Society*, 99: 315-331.

Hu, X., J.P. Rosenfeld, and G.V. Bodenhausen. 2012. Combating Automatic Autobiographical Associations: The Effect of Instruction and Training in Strategically Concealing Information in the Autobiographical Implicit Association test. *Psychological Science*, DOI: 10.1177/0956797612443834, 1-7.

Kamtekar, R. 2004. Situationism and Virtue Ethics on the Content of Our Character. *Ethics*, 114: 458-491.

Kihlstrom, J. 2008. The Automaticity Juggernaut—or Are We Automaton After All?. In: J. Baer, J. Kaufman, and R. Baumeister, eds. *Are We Free? Psychology and Free Will*. New York: Oxford University Press.

Levy, N. 2012. Consciousness, Implicit Attitudes, and Moral Responsibility. *Noûs*, DOI: 10.1111/j.1468-0068.2011.00853.x, 1-22.

Libet, B. 1985. Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action. *Behavioral and Brain Sciences*, 8: 529-66.

——— 1999. Do We Have Free Will?. *Journal of Consciousness Studies*, 6: 47-57.

——— 2004. *Mind Time*. Cambridge, Mass.: Harvard University Press.

Lieberman, M. 2005. Principles, Processes, and Puzzles of Social Cognition. *NeuroImage*, 28: 746-756.

Mele, A. 2009. *Effective Intentions*. Oxford: Oxford University Press.

——— 2012. *Backsliding*. Oxford: Oxford University Press.

——— n.d. Free Will and Substance Dualism: The Real Scientific Threat to Free Will?. In: W. Sinnott-Armstrong, ed. *Moral Psychology, Volume 4: Free Will and Responsibility*. Cambridge, Mass: MIT Press.

Milgram, S. 1963. Behavioral Study of Obedience. *The Journal of Abnormal and Social Psychology*, 67: 371-78.

——— 1965. Some Conditions of Obedience and Disobedience to Authority. *Human Relations*, 18: 57-76.

——— 1974. *Obedience to Authority*. New York: Harper & Row.

Rigdon, M., K. Ishii, M. Watabe, and S. Kitayama. 2009. Minimal Social Cues in the Dictator Game. *Journal of Economic Psychology*, 30: 358-367.

Newell, B. and D. Shanks. n.d. Unconscious Influences on Decision Making: A Critical Review. *Behavioral and Brain Sciences*.

Sabini, J. and M. Silver. 2005. Lack of Character? Situationism Critiqued. *Ethics*, 115: 535-562.

Sreenivasan, G. 2002. Errors about Errors: Virtue Theory and Trait Attribution. *Mind*, 111: 47-68.

Stewart, B. and K. Payne. 2008. Bringing Automatic Stereotyping Under Control: Implementation Intentions as Efficient Means of Thought Control. *Personality and Social Psychology Bulletin*, 34: 1332-1345.

Tong, F., K. Nakayama, M. Moscovitch, O. Wienrib, and N. Kanwisher. 2000. Response Properties of the Human Fusiform Face Area. *Cognitive Neuropsychology*, 17: 257-279.

Uhlmann, E. and G. Cohen. 2005. Constructed Criteria: Redefining Merit to Justify Discrimination. *Psychological Science*, 16: 474-480.

Webb, T., P. Sheeran, and A. Luszczynska. 2009. Planning to Break Unwanted Habits: Habit Strength Moderates Implementation Effects on Behavior Change. *British Journal of Social Psychology*, 48: 507-523.

Wegner, D. 2002. *The Illusion of Conscious Will*. Cambridge, Mass. MIT Press.

——— 2004. Précis of 'The Illusion of Conscious Will'. *Behavioral and Brain Sciences*, 27: 649-659.

——— 2008. Self is Magic. In: J. Baer, J. Kaufman, and R. Baumeister, eds. *Are We Free? Psychology and Free Will*. New York: Oxford University Press.

Zimbardo, P. n.d. "Stanford Prison Experiment": <http://www.prisonexp.org/>.

Zimbardo, P., C. Haney, W. Banks, and D. Jaffe. 1973. The Mind Is a Formidable Jailer: A Pirandellian Prison. *The New York Times Magazine*, Section 6 (April 8: 38-60).<http://www.prisonexp.org/>.<http://www.prisonexp.org/>.