Utilitas (2020), **32**, 50–69
doi:10.1017/S0953820819000335

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Civilian Immunity Without the Doctrine of Double Effect

Yitzhak Benbaji* and Susanne Burri*

Tel Aviv University and London School of Economics and Political Science
*Corresponding author. E-mail: ybenbaji@tauex.tau.ac.il s.burri@lse.ac.uk

(First published online 9 September 2019)

Abstract
Civilian Immunity ('Immunity') is the legal and moral protection that civilians enjoy against the effects of hostilities under the laws of armed conflict and according to the ethics of killing in war. Immunity specifies different permissibility conditions for *directly targeting civilians* on the one hand, and for *harming civilians incidentally* on the other hand. Immunity is standardly defended by appeal to the Doctrine of Double Effect (DDE). We show that Immunity's prohibitive stance towards targeting civilians directly, and its more permissive stance towards harming them incidentally, can be defended without appealing to the DDE if agents suffer from *overconfidence*. Overconfidence is a cognitive bias that affects agents who are required to make decisions in the presence of significant uncertainty.

I. Introduction

Civilian Immunity (hereinafter 'Immunity') is the legal and moral protection that civilians enjoy against the effects of hostilities under the laws of armed conflict[1] and according to the ethics of war.[2] International law characterizes the protection as follows:

[1]We do not aim to defend the *legal* contours of Civilian Immunity in this article. Famously, international law does not distinguish between combatants who are fighting for a just cause, and combatants who are fighting for an unjust cause. Our concern in this article is only with *just* combatants, and the constraints that they are under when fighting for a just cause. We set aside the legal and moral status of combatants fighting for an unjust cause.

[2]According to the ethics of killing in war, it is relatively uncontroversial that Immunity applies to at least those civilians who do not share in the moral responsibility for an unjust war that their country is fighting. It is contested whether Immunity applies to other civilians as well (for an argument against, see Jeff McMahan, 'The Ethics of Killing in War', *Ethics* 114 (2004), pp. 693–733, at 725–9; for an argument in favour, see Henry Shue, 'Do We Need a "Morality of War"?', *Just and Unjust Warriors*, ed. David Rodin and Henry Shue (Oxford, 2008), pp. 87–111). It is not the purpose of our article to delimit precisely which civilians – or which individuals more generally – are morally protected by Immunity. We instead aim to investigate how Immunity can be justified in non-controversial cases, i.e. in cases where the physical integrity of non-threatening, non-responsible individuals is at stake. For ease of presentation, we label this set of individuals 'civilians'. See also n. 4.

© Cambridge University Press 2019

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

> In order to ensure respect for and protection of the civilian population and civilian objects, the parties to the conflict shall at all times distinguish between the civilian population and combatants and … accordingly shall direct their operations only against military objectives. . . . The civilian population as such, as well as individual civilians, shall not be the object of attack.[3]

As it is usually understood, Immunity prohibits directly targeting civilians even if doing so would further some important military objective.[4]

The rules on incidentally harming civilians are more permissive. If harm to civilians is unsought, attacks on military targets can sometimes be permissible, even when they will predictably lead to civilian casualties.[5] More precisely, combatants may incidentally harm civilians in the pursuit of a military advantage whenever they take care to minimize such harm, and whenever the harm that is expected is not 'excessive in relation to the concrete and direct military advantage anticipated'.[6] While Immunity thus *invariably* protects civilians against *being targeted*, it protects them against *incidental* harm only in cases where such harm is *excessive*, or where no adequate effort is undertaken to *minimize* the harm.

The following two hypothetical cases[7] illustrate the basic distinction that Immunity draws:

> *Terror Bombing*: A terror bomber blows up the house of an enemy village elder and his family. The village elder is much beloved in his community, and the terror bomber targets his house to demoralize the enemy. She knows that her attack will kill ten people, and that it will create a welfare loss of −500. She also knows that the military advantage secured by the demoralization will achieve welfare gains worth 1,000.

> *Tactical Bombing*: A tactical bomber bombs a munitions factory on enemy territory to destroy its productive capacity. The tactical bomber knows that bombing the munitions factory will kill ten enemy civilians – a beloved village elder and his family – who live near the factory, thus creating a welfare loss of −500. She also knows that the military advantage secured by the destruction of the factory will achieve welfare gains worth 1,000.

Immunity categorically prohibits the bombing in *Terror Bombing*. By contrast, it prohibits the bombing in *Tactical Bombing* only if killing the ten villagers is excessive or if harm to them was not minimized. From a welfarist perspective – that is, from a perspective where only the welfare that comes to different individuals is thought to have ultimate moral significance – these differential verdicts are puzzling. After all, the two cases share the same welfarist profile, and it is stipulated in both cases that the

[3]*Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts*, 8 June 1977 (AP I), articles 48 and 51(2).

[4]The prohibition on targeting civilians as laid out in AP I is not absolute. According to article 51(3) of AP I, civilians who are 'directly participating in hostilities' lose their right not to be targeted. See also n. 2.

[5]AP I, article 51(5).

[6]AP I, article 57(2.a.ii–iii).

[7]In the philosophical literature, Jonathan Bennett is one of the first to mention such a set of cases. See Jonathan Bennett, *Morality and Consequences: The Tanner Lectures on Human Values* (Cambridge, 1980), p. 95.

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

agent has perfect foresight and knows for certain what will be the consequences of her actions. It therefore seems clear that only a non-welfarist principle can help us make sense of Immunity.

In the ethics of war literature, a non-welfarist principle called the *Doctrine of Double Effect* (DDE) is standardly adduced in attempts to explain and justify Immunity. According to the DDE, it is morally worse to inflict harm on innocent individuals when the harm is *intended* rather than *merely foreseen*. The DDE thus implies that it can sometimes be morally permissible to bring about a sufficiently important moral good at the cost of inflicting some merely foreseen harm, but not at the cost of inflicting the same harm intentionally. If the pilot in *Terror Bombing* bombs the village elder's house, it is plausible that she kills her victims intentionally, as she deliberately targets them. In *Tactical Bombing*, by contrast, if the pilot targets the munitions factory, it is plausible that she merely foresees that bombing the factory will kill civilians. The non-welfarist distinction that the DDE draws thus seems well aligned, or possibly even to coincide, with the distinction that Immunity draws. The problem with this explanation is that it is difficult to explain what makes the non-welfarist distinction between intended and merely foreseen harm morally significant.

In this article, we aim to show that when an agent's deliberative conditions are non-ideal – as is plausibly frequently the case in war – *welfarist* considerations have the potential to validate the seemingly *non-welfarist* distinction that Immunity draws. More specifically, we argue that appealing to the DDE is not necessary to defend Immunity's prohibitive stance towards targeting civilians directly, and its more permissive stance towards harming them incidentally, when agents suffer from a cognitive bias called *overconfidence*.

The article is structured as follows. In section II, we motivate our project. Section III advances an analysis of welfarist deliberation under circumstances of *ideal uncertainty*. Under *ideal uncertainty*, a rational agent can carefully deliberate about all the possible consequences of her alternatives for action. We show that under ideal uncertainty, welfarism implies that direct killings of civilians tend to be harder to justify than merely incidental killings. Section IV shows that when we give up the assumption of ideal deliberative circumstances and move to conditions that we refer to as *normal uncertainty*, welfarism not only accords with the basic distinction that Immunity draws; it moreover has the potential to justify abiding by Immunity's more specific prescriptions. Section V concludes.

II. Why attempt a welfarist defence of Immunity?

On the assumption that agents know for certain what will be the consequences of their actions, there is no welfarist difference between *Terror Bombing* and *Tactical Bombing*, and the fact that Immunity distinguishes between the two cases appears unmotivated from a welfarist perspective. This makes it seem obvious that an appropriate defence of Immunity must be non-welfarist in character. One might worry that attempting a welfarist defence of Immunity risks distorting the principle, and threatens to downgrade it to a mere heuristic.

We think that these worries are justified, and need to be guarded against. For four reasons, we nevertheless believe that it is important and worthwhile to investigate what a welfarist defence of Immunity might look like.

First, moral philosophers are frequently friendly to the idea that under imperfect deliberative circumstances, deontological principles might serve as useful heuristics

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

from a welfarist point of view. But it is rarely investigated in any detail what an agent's deliberative circumstances would have to look like for the agent actually to be warranted, from a welfarist perspective, in relying on a particular deontological principle.[8] As we hope to clarify in this article, this is not a straightforward task. To get a welfarist defence of Immunity off the ground, we need to introduce a number of empirical assumptions. While we defend these assumptions as *frequently justified in the context of war*, difficult steps in the argument remain. We discuss these difficulties as they arise, thus highlighting that a deontological principle such as Immunity receives support from a welfarist perspective only in a limited set of deliberative contexts.

A second reason in favour of working out a welfarist defence of Immunity is as follows. Many individuals endorse welfarist moral frameworks. They are committed to the claim that if two alternatives share the same welfarist profile, then this renders them morally equivalent. If we can explain to such individuals that there are circumstances under which an agent has reason to abide by the strictures of Immunity even if welfarism is correct, Immunity's reach is extended.

Third, the standard non-welfarist defence of Immunity rests on shaky grounds. Despite its intuitive appeal, the DDE remains a controversial doctrine even among those who are open to non-welfarist justifications. For one thing, many doubt that there exists a meaningful conceptual distinction between intended and merely foreseen harm that is capable of separating cases such as *Terror Bombing* from cases such as *Tactical Bombing*. As Jonathan Bennett has argued, most harms that appear intentionally inflicted because they seem to be *instrumentally necessary* to the achievement of an end turn out to be in principle dispensable on closer inspection.[9] Consider the situation of the terror bomber. To achieve her aim of demoralizing the enemy, it suffices that she brings about a situation where the village elder and his family *appear dead* until the end of the war.[10] A well-intentioned terror bomber might thus narrowly intend only to make her victims appear dead, while foreseeing that, under the circumstances, making them appear dead will also have the effect of killing them. Once we describe a well-intentioned terror bomber in this manner, it seems that her attitude towards her victims becomes indistinguishable from the attitude of a well-intentioned tactical bomber.[11]

Some philosophers have argued that the traditional DDE should be replaced with a modified principle. Their basic suggestion is that if an action will foreseeably cause

[8]For relevant exceptions to this rule, see Richard Brandt, 'Utilitarianism and the Rules of War', *Philosophy & Public Affairs* 1 (2, 1971), pp. 145–65, and William H. Shaw, *Utilitarianism and the Ethics of War* (London and New York, 2016). Brandt's key claim is that from a utilitarian point of view, we have reason to implement necessity and proportionality constraints on the pursuit of military objectives. We discuss Shaw's views in section IV below.

[9]See Jonathan Bennett, *The Act Itself* (Oxford, 1998), ch. 11.

[10]Bennett, *Morality and Consequences*, p. 111.

[11]A standard reply to Bennett's objection contends that an agent intends harm if the harm is *sufficiently close* to the agent's goal, or to a necessary means to the agent's goal. For an early such argument, see Philippa Foot, 'The Problem of Abortion and the Doctrine of the Double Effect', in *Virtues and Vices* (Berkeley, 1978), pp. 19–32, at 21–2. For more recent arguments, see e.g. William FitzPatrick, 'The Intend/Foresee Distinction and the Problem of "Closeness" ', *Philosophical Studies* 128 (2006), pp. 585–617; Ralph Wedgwood, 'Defending Double Effect', *Ratio* (new series) 24 (2011), pp. 384–401; Alison Hills, 'Intentions, Foreseen Consequences and the Doctrine of Double Effect', *Philosophical Studies* 133 (2007), pp. 257–83. These approaches face many difficulties, see e.g. See Dana K. Nelkin and Samuel C. Rickless, 'So Close, Yet So Far: Why Solutions to the Closeness Problem for the Doctrine of Double Effect Fall Short', *Noûs* 49 (2015), pp. 376–409. Compare Victor Tadros, 'Wrongful Intentions Without Closeness', *Philosophy & Public Affairs* 43 (2015), pp. 52–74, at 59–60.

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

harm to an innocent person, then this action is especially hard to justify if the agent involves her victims in the pursuit of aims that the victim neither shares, nor is morally obliged to share.[12] In particular, according to the *Means Principle*, it is especially morally problematic harmfully to use an innocent person as a means to the promotion of one's ends. Suggestions along these lines avoid the closeness problem, but they have been criticized for having counterintuitive implications.[13]

We believe that as long as Immunity rests on shaky non-welfarist grounds, it is worthwhile to put it on a more secure footing by showing how welfarist considerations are able to support it. Further, welfarist considerations are of interest not only to 'pure welfarists' who endorse some welfarist moral theory. Instead, they are of interest also to 'impure welfarists' who believe that welfarist considerations have some role to play within a more complex moral theory, where they need to be balanced against, and will sometimes be trumped by, other morally relevant considerations. For such impure welfarists, a welfarist defence of Immunity helps underscore the principle's practical relevance, which makes it of interest even as the search for a compelling non-welfarist justification of the principle continues.

Finally, there exists a mode of inflicting harm – following Judith Thomson, we refer to it as *riding roughshod harming*[14] – where Immunity's rulings are unclear. Riding roughshod harming is underexplored in the literature, and in informal discussions, we have registered a disagreement among law scholars about the extent to which Immunity protects civilians against it. When an agent inflicts riding roughshod harm, she foreseeably harms a victim in the lead-up to securing a greater good. Consider the following case:

> *Riding Roughshod Bombing*: Tanks are on their way to bomb an enemy munitions factory. They will only get to the munitions factory on time if they run over an enemy house with ten civilians locked up inside it, creating a welfare loss of −500. The military advantage that is achieved by destroying the factory secures welfare gains worth 1,000.

Intuitively, it seems clear that running over the civilians in *Riding Roughshod Bombing* is harder to justify than bombing the civilians in *Tactical Bombing*. Alas, neither the DDE nor the Means Principle can explain why this is so. The tank drivers do not intend to kill any civilians; they merely foresee that rushing to the munitions factory will result in such killings. To date and to our knowledge, no other non-welfarist principle has been put forward that can satisfactorily explain the moral difference between *Riding Roughshod Bombing* and *Tactical Bombing*. The unresolved status of roughshod harm is another reason to explore how welfarist considerations might illuminate Immunity.

[12]Such a proposal was first articulated by Warren Quinn in his 'Actions, Intentions, and Consequences: The Doctrine of Double Effect', *Philosophy and Public Affairs* 18 (1989), pp. 334–51. Versions of Quinn's proposal have recently been defended by Tadros, 'Wrongful Intentions', and by Dana K. Nelkin and Samuel C. Rickless, 'Three Cheers for Double Effect', *Philosophy and Phenomenological Research* 89 (2014), pp. 125–58.

[13]See e.g. FitzPatrick, 'The Intend/Foresee Distinction'; Thomas A. Cavanaugh, *Double-Effect Reasoning: Doing Good and Avoiding Evil* (Oxford, 2006), pp. 59–66; Frances Kamm, *Intricate Ethics: Rights, Responsibilities, and Permissible Harm* (Oxford, 2007), pp. 82–9. For a reply, see Nelkin and Rickless, 'Three Cheers'.

[14]Judith Thomson, 'Self-Defense', *Philosophy & Public Affairs* 20 (1991), pp. 283–310, at 290.

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

In the remainder of this article, we replace the artificial assumption that agents have perfect foresight with more realistic scenarios that involve *deliberative uncertainty*. We argue that, under deliberative uncertainty, welfarist considerations validate the seemingly non-welfarist distinction that Immunity draws, and they moreover provide support for the intuitive judgement that roughshod harm is harder to justify than other merely foreseeable harm.

III. Welfarism under uncertainty

III.1. Welfarist moral theories

In this article, we use the term *welfarism* specifically to refer to maximizing act consequentialist moral theories whose axiologies satisfy a *welfarist constraint*. An axiology is a moral betterness ranking of states of affairs. On our understanding of the term, an axiology satisfies a *welfarist constraint* just in case it measures the goodness or badness of a state of affairs in terms of a function that takes as its input only the welfare that individuals enjoy in the state of affairs.[15]

Due to their maximizing act consequentialist nature, welfarist moral theories define the objective rightness of an action as follows:

> *Welfarist criterion for the rightness of an action*: An action is objectively right just in case it leads to an outcome such that no other available action would lead to a morally better outcome in welfarist terms.

In the *Bombing* cases as we have introduced them, it is stipulated that agents know for certain what will be the consequences of their actions. Under such circumstances of perfect foresight, the welfarist criterion for right action easily translates into the following action-guiding principle:

> *Action-guiding principle for agents with perfect foresight*: From your set of available actions, always perform an action that will lead to one of the best possible outcomes.

As *Terror Bombing*, *Tactical Bombing*, and *Riding Roughshod Bombing* all share the same welfarist profile, an agent with perfect foresight should treat them as indistinguishable from a welfarist point of view, and should always choose to secure the military advantage, thus creating a net welfare gain of $1{,}000 - 500 = 500$.

III.2. Ideal uncertainty

In real life, while we frequently have a good idea of the *likely* consequences of our actions, we usually also know that things could turn out differently than we expect. In this article, we use the term *ideal uncertainty* to refer to the limited foresight that tends to characterize our epistemic situation even when our deliberative circumstances are otherwise ideal. More precisely, we use the term *ideal uncertainty* for circumstances where an agent (i) knows her alternatives for action, (ii) knows how to evaluate the

[15]We assume that welfare is interpersonally comparable and measurable on a cardinal scale. This makes it reasonable to assume that under ideal uncertainty, agents who endorse a welfarist moral theory should aim at maximizing expected value (see sec. III.2 and, in particular, note 17).

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

different outcomes that her actions might bring about, and where the agent (iii) has the resources to assign unbiased probability estimates to the occurrence of possible outcomes conditional on her performance of alternative actions. Ideally uncertain agents reason in an unbiased manner and make full use of the extensive information they have access to.

How should an ideally uncertain agent deliberate if she believes that some welfarist moral theory is correct? Because she is unsure what consequences her actions will have, she cannot simply decide to bring about the best possible outcome. It seems reasonable to assume that an ideally uncertain agent should make use of all the available information, and should therefore take into account both the *values* of the states of affairs that her actions might bring about, as well as the *probabilities* with which different states of affairs might occur. A straightforward way of doing this is by calculating the expected value of each action, and then performing that action which has the highest expected value. When the values of states of affairs can be estimated cardinally – as we suppose is the case for welfarism (see note 15) – this deliberative strategy is simple and intuitively attractive. We therefore propose the following deliberative principle for circumstances of ideal uncertainty:

> *Action-guiding principle for deliberative circumstances of ideal uncertainty*: From your set of available actions at the time of choice (t_{act}),[16] always perform an action which has maximal expected value as estimated at t_{act}, where the expected value of an action is the sum of the probability-weighted welfare values of the outcomes that the action might bring about.[17]

When we introduce ideal uncertainty to *Tactical Bombing* and *Terror Bombing*, a systematic probability gap that is absent in *Tactical Bombing* becomes apparent in *Terror Bombing*. This difference between the two cases means that under ideal uncertainty, welfarism supports the seemingly non-welfarist claim that some ways of bringing about harm to innocent individuals are harder to justify than others.

Consider *Terror Bombing* under ideal uncertainty. If you are the terror bomber, and your foresight is limited in the usual way, you will not know for certain that killing the village elder and his family will achieve the desired effect of demoralizing the enemy. It is, for example, quite possible that the enemy will be angered instead of demoralized. If you consider the effects of terror bombing under ideal uncertainty, you thus reasonably assign a higher probability to the occurrence of the anticipated *bad effect* (dead civilians) than to the occurrence of the anticipated *good effect* (demoralization of the enemy).

The just described probability gap arises because, from the agent's point of view, killing the civilians is something that she needs to get done as *part of her plan* of demoralizing the enemy. It is an important step towards demoralizing the enemy, but, at the same time, dead civilians are merely a *link* in a longer causal chain that the agent *projects* will lead to the demoralization of the enemy. When an ideally uncertain deliberator

[16]This terminology will be introduced in more detail in section IV below.

[17]This principle is one of a range of reasonable principles to guide welfarist deliberation under ideal uncertainty. More specifically, we concede that agents needn't be risk-neutral, but may reasonably be at least moderately risk-averse. The results that we discuss obtain whether agents are risk-neutral or risk-averse, and hence we do not discuss admissible risk aversion in the main text. Note that depending on how the cardinalization of welfare values is obtained, attitudes towards risk might already be embedded in the numbers that describe the values of outcomes.

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

recognizes that an expected consequence C of one of her available actions is causally prior to another expected consequence C′, she reasonably takes this fact into account by assigning a higher probability to C than to C′ actually occurring conditional on her performing the action. The further causally removed an expected consequence of an action, the worse an ideally uncertain agent's epistemic access to it, and the lower should be her credence in its occurrence conditional on her performing the action.

That our epistemic access to the consequences of our actions worsens as these consequences become causally more remote is an aspect of the very nature of human cognition. Consider a basic action such as moving your right index finger. If you are thinking about doing this, you can normally be confident that if you choose to move your finger, you will in fact do so. Your epistemic access to the consequences of your basic actions is generally excellent because these consequences are usually almost entirely under your control. The same tends not to be true for causally more removed effects. If you move your right index finger to pull a trigger and shoot an attacker, whether your plan will succeed depends on many factors that are not directly under your control. Your plan might fail because your gun is jammed, or because a bird crosses the path of your bullet, or because the attacker dodges your bullet by ducking. Thanks to what you know about the world, you will often be able to predict that one of your actions is *likely* to help bring about some causally removed consequence, but you cannot usually be sure.

In *Terror Bombing*, the probability gap between the two anticipated effects of bombing the village elder's house lowers the action's expected welfare value compared to a situation of perfect foresight. If the probability gap between the two anticipated effects is large, or if the welfare value of the good effect exceeds the disvalue of the bad effect by only a small margin, an ideally uncertain welfarist deliberator may do well to abstain from an option that an agent with perfect foresight would rightly pursue.

For an illustrative example of the probability gap, see Table 1. In this example, we have set the welfarist value of the military advantage at 1,000, the welfarist disvalue of the harm to civilians at −500, and the probability that the enemy will be demoralized if the village elder's house is bombed at 0.6.[18] This yields an expected welfare value of 100 for the option of bombing the house, and an expected welfare value of 0 for abstaining from the bombing. This is much lower than the welfare value of *Terror Bombing* under perfect foresight, which coincides with the welfare value of 500 of outcome 2 in Table 1.

In *Tactical Bombing*, no comparable systematic probability gap obtains, as the underlying causal structure of the case is different. If you are the tactical bomber, and your foresight is limited in the usual way, you should not conceptualize bombing the village as merely a *link* in a longer causal chain that you *project* will lead to the destruction of the munitions factory. Instead of a step towards destroying the factory, harming the civilians in *Tactical Bombing* is a 'noncausal flip side'[19] of destroying the factory, as the same event that you reasonably expect will destroy the factory – the blast of the dropped bombs – is also the event that you reasonably expect will kill the civilians.

It is possible that, in particular cases, the anticipated bad effect of tactically bombing the munitions factory is more probable than the anticipated good effect, for example because an ideally uncertain deliberator is doubtful that a targeted structure really serves

[18]For simplicity, we here assume that 'being demoralized' is a binary notion.

[19]Kamm, *Intricate Ethics*, p. 141.

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

Table 1. *Terror Bombing* under ideal uncertainty

			Possible outcomes and their probabilities of occurring depending on the action taken			
		Expected value of action	(1) Military advantage achieved (value = 1,000), civilians not harmed	(2) Military advantage achieved (value = 1,000), civilians harmed (value = −500)	(3) Military advantage not achieved, civilians not harmed	(4) Military advantage not achieved, civilians harmed (value = −500)
Action	Bomb	0.6*500 + 0.4* (−500) = 100	Probability = 0	Probability = 0.6	Probability = 0	Probability = 0.4
	Do not bomb	1*0 = 0	Probability = 0	Probability = 0	Probability = 1	Probability = 0

as a munitions factory. But this type of uncertainty can also go the other way: an ideally uncertain deliberator might for example have some reason to think that the house next to the munitions factory has been evacuated. It follows that this type of possible uncertainty does not introduce a *systematic* probability gap that makes the occurrence of the anticipated bad effect *consistently* more likely than the occurrence of the anticipated good effect. Nevertheless, depending on the circumstances, the size of such non-systematic probability gaps might sometimes dwarf the size of a systematic probability gap, thus swamping the latter's deliberative significance. Our conclusion is therefore modest: under ideal uncertainty, welfarism implies that *other things equal*, welfare losses are harder to justify if they occur causally prior to offsetting welfare gains.

Table 2 provides a numerical illustration of *Tactical Bombing* under ideal uncertainty. In this table, we account for the just mentioned possibility of unsystematic probability gaps with probabilities of 0.1 for the first and fourth possible outcomes. Assuming the same welfare values as in Table 1, the expected welfare value of bombing the factory is 450, which is substantially higher than the expected welfare value of 100 associated with terror bombing.

In summary, cases where harm is intended in the pursuit of some greater good share a *causal structure* that need not characterize cases of unintentional harming, namely a structure where the *occurrence of the harm is causally prior to the occurrence of the greater good*. This causal structure means that harm inflictions that Immunity and the DDE flag as especially morally problematic are cases where typically there is a real risk that trying to achieve some worthy aim ends in disaster. That is, in cases that Immunity and the DDE flag as especially morally problematic, there will usually be a non-negligible chance that while the anticipated *bad* effect is brought about, the anticipated *good* effect fails to materialize.

Importantly, this welfarist analysis of the *Bombing* cases under ideal uncertainty groups *Riding Roughshod Bombing* with *Terror Bombing*, thus distinguishing it from *Tactical Bombing*. Riding roughshod cases are cases where harm is inflicted unintentionally, but *causally prior* to the securing of a greater good. It follows that an ideally uncertain deliberator reasonably assigns a higher probability to the occurrence of roughshod harm than to the achievement of the intended good effect.

Someone might object to this analysis with the following observation: in the presentation of *Tactical Bombing* in Table 2, we have stipulated a welfarist value of 1,000 for the military advantage secured through the destruction of the factory. But this value is itself merely an *estimate* of what the military advantage is thought *eventually* to help achieve, for example by saving lives through shortening the war. It follows that the harm inflicted on civilians in *Tactical Bombing* is causally prior to the good that it is ultimately expected to help secure. And this in turn would seem to imply that, from a welfarist perspective and under ideal uncertainty, there is no relevant difference between *Terror Bombing* and *Tactical Bombing* after all. Instead, in both cases, harm to enemy civilians is causally prior to the greater good that is expected ultimately to redeem this harm.

We concede that, from a welfarist point of view, the moral value of a military advantage needs to be calculated in terms of the beneficial welfare effects that it helps secure, which makes it very difficult to estimate the value of *any* military advantage. Having said that, the difference between cases such as *Terror Bombing* and cases such as *Tactical Bombing* that we have argued for in this section nevertheless remains intact. In *Terror Bombing*, it is unclear that bombing the civilians will secure the military advantage of demoralizing the enemy, *however the value of this advantage is best*

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

Table 2. *Tactical Bombing* under ideal uncertainty

			Possible outcomes and their probabilities of occurring depending on the action taken			
		Expected value of action	(1) Military advantage achieved (value = 1,000), civilians not harmed	(2) Military advantage achieved (value = 1,000), civilians harmed (value = −500)	(3) Military advantage not achieved, civilians not harmed	(4) Military advantage not achieved, civilians harmed (value = −500)
Action	Bomb	0.1*1,000 + 0.8*500 + 0.1* (−500) = 450	Probability = 0.1	Probability = 0.8	Probability = 0	Probability = 0.1
	Do not bomb	1 * 0 = 0	Probability = 0	Probability = 0	Probability = 1	Probability = 0

estimated. In *Tactical Bombing*, by contrast, the agent can be confident that bombing the civilians will secure the military advantage of destroying the munitions factory, even if the further question of how best to estimate the value of this advantage remains. To put the same point differently, note that in the context of war and under ideal uncertainty, achieving a military advantage is reasonably conceptualized as an important *milestone in the securing of welfare gains* that winning the war is ultimately thought to realize. If such a milestone comes with associated welfare costs that are projected to be causally concurrent with, or causally posterior to, the achievement of the milestone, this is preferable to comparable costs occurring causally prior to the milestone – for precisely the reasons that we have identified in this section.

Of course, we do not mean to suggest that the probability gap that we have identified will *always* play an *important* role in deliberation under ideal uncertainty. Where it is present, it may sometimes be small. If soldiers kill a child in front of the child's father to incapacitate the father, the soldiers might reasonably be almost certain that they will achieve their intended effect. Further, and as already mentioned, where the probability gap is absent, this does not guarantee that an expected good effect is at least as likely as an expected bad effect. In a case such as *Tactical Bombing*, a tactical bomber might for example have reason to suspect that a targeted structure no longer serves as a munitions factory. If she decides to bomb the factory, it might then be more likely that she will kill civilians than that she will secure a military advantage. Our modest ambition in this section was simply to argue that under ideal uncertainty, the probability gap that we have identified is of *some* significance from a welfarist point of view. In the next section, we discuss deliberative conditions under which its significance is amplified.

IV. Welfarism under normal uncertainty

Welfarism under ideal uncertainty implies that, other things equal, ways of harming civilians that Immunity rules out as impermissible are harder to justify than ways of harming civilians that Immunity treats more permissively. Due to their causal structure, cases such as *Terror Bombing* come with a systematic probability gap under ideal uncertainty that lowers their expected value compared to otherwise similar cases whose causal structure resembles *Tactical Bombing*. But this probability gap does not imply that there is a general prohibition on inflicting the type of harm that Immunity prohibits. Instead, an ideally uncertain agent is advised always to perform that action which maximizes expected value.

Things change once we move from ideal uncertainty to less favourable deliberative circumstances. In what follows, we argue that under imperfect deliberative circumstances, welfarism has the potential to support Immunity's injunctions.

IV.1. Normal uncertainty in war

Normal uncertainty obtains when a deliberating agent is unable to evaluate in a comprehensive and unbiased manner her alternatives for action. More precisely, normal uncertainty obtains when at least one of the following *fails* to hold: (i) the agent knows what her alternatives for action are; (ii) the agent knows how to evaluate the different outcomes that her actions might bring about; (iii) the agent is able to assign unbiased probability estimates to the occurrence of possible outcomes conditional on her performance of alternative actions. The factors that induce a shift from ideal to normal uncertainty include:

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

- *Time pressure.* An agent must make up her mind fast; if she does not act quickly, some of her alternatives for action may no longer be available.
- *Incomplete empirical information.* An agent has to decide how to act when it is neither entirely clear what her alternatives for action are, nor which outcomes choosing the different alternatives would entail.
- *Incomplete evaluative information.* An agent must decide how to act in circumstances where her objectives are underspecified, or where she does not know how to evaluate different possible outcomes.
- *Cognitive overload.* An agent has to decide how to act when many things are on her mind.
- *Distress.* An agent has to decide how to act when much is at stake.

Empirical research shows that we are prone to deliberate in biased ways when even just some of these factors are present, as they lead us to disregard some of the available information, and to rely on invalid inferences.[20] On the plus side, reasoning in this 'fast and frugal'[21] manner allows us to cope with complex and challenging circumstances. On the minus side, when we reason fast and frugally, the quality of our decision-making is compromised compared to a situation where we deploy our reasoning capacities in a slower, more deliberate, and maximally systematic manner.

In war, deliberative circumstances are frequently far from ideal. William Shaw argues that, since utilitarianism permits defensive wars, and since such wars cannot be fought without killing civilians, the distinction that Immunity draws between incidental and intentional killings is justified from a utilitarian point of view.[22] He contends that intentionally targeting civilians is highly unlikely to be a welfare-maximizing strategy, and that the non-ideal deliberative circumstances in war render it close to impossible to identify possible exceptions to this general rule.[23] For Shaw, this makes it optimal, from a welfarist point of view, that all combatants should cultivate a deep-seated aversion towards intentionally targeting civilians, and that they should refrain from such targeting at all times.[24]

While we are sympathetic to Shaw's line of reasoning, we believe that more needs to be said *why* targeting civilians under non-ideal deliberative circumstances is unlikely to be a welfare-maximizing strategy. In this section, we provide one possible explanation. More specifically, we argue that if imperfectly deliberating agents suffer from a bias called *overconfidence*, then the probability gap that we have identified in section III may frequently lead such biased agents to overestimate vastly the expected welfare effects of intentionally targeting civilians.

Overconfidence is a pervasive tendency of human beings to overestimate the probability with which they will succeed in their endeavours, essentially because they tend to

[20]See e.g. Daniel Kahneman, *Thinking: Fast and Slow* (London, 2012); Daniel Kahneman and Dan Lovallo, 'Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking', *Management Science* 39 (1993), pp. 17–31, esp. at 25; Marianne Bertrand, Dolly Chugh, and Sendhil Mullainathan, 'Implicit Discrimination', *AEA Papers and Proceedings* 95.2 (2005), pp. 94–8, esp. at 96.

[21]Gerd Gigerenzer and Daniel G. Goldstein, 'Reasoning the Fast and Frugal Way: Models of Bounded Rationality', *Psychological Review* 103 (1996), pp. 650–69.

[22]Shaw, *Utilitarianism and the Ethics of War*, pp. 120–40. See also William H. Shaw, 'Utilitarianism and the Ethics of War', *The Cambridge Guide to Utilitarianism*, ed. Ben Eggleston and Dale E. Miller (Cambridge, 2014), pp. 303–24, esp. at 321–3.

[23]Shaw, *Utilitarianism and the Ethics of War*, pp. 130–1.

[24]Shaw, *Utilitarianism and the Ethics of War*, pp. 130–1.

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

'exaggerate their control over events, and the importance of the skills and resources they possess in ensuring desirable outcomes'.[25] Overconfidence is a bias that it is difficult to correct; even when we are aware of its existence, the bias tends to persist, as it is resistant to 'logic, decomposition, or the use of training and tools'.[26] Studies show that overconfidence crops up whenever *complex quantitative estimates* need to be made, for example in defence, legal, financial, managerial and engineering decisions.[27] Based on case studies ranging from World War I to the Iraq War, Dominic D. P. Johnson argues that overconfidence is a pervasive feature of decision-making in war.[28]

IV.2. The probability gap under normal uncertainty

Imagine a normally uncertain agent in war who is presented with an opportunity to pursue some military advantage in a way that threatens harm to enemy civilians. If the agent is overconfident, she will have little doubt that making use of the opportunity will help her secure the military advantage in question. To give a numerical example, let us assume that an overconfident agent assigns a 90% probability to succeeding with a particular mission, when a careful and unbiased evaluation of the situation would indicate that the probability of success is no more than 50%. Now, crucially, if the harm to enemy civilians is *causally prior* to the securing of the military advantage, this overconfidence spells trouble, as it will tend to obscure probability gaps. To see what we mean, note that the higher the probability that an agent assigns to succeeding, the less deliberative space she has for recognizing a probability gap between the causally prior harm and the military advantage. This is so because the occurrence of a causally prior effect has to be seen as equally or more likely than the occurrence of a posterior effect, so that an agent who assigns a *high* probability to a *posterior* effect is left with little wiggle room to assign *differential* probabilities to the two effects. Suppose that in the numerical example, the probability of the causally prior harm is 95%, and that the probability gap between the occurrence of the harm vs the securing of the military advantage is thus 45%. Once the agent decides that her chances of success are roughly 90%, the largest gap that she can still recognize is one of 10%, which she does in case she assigns a 100% chance to the probability of harm. To put it in a catchy phrase, overconfident agents will frequently struggle to 'mind the gap'.

The situation is different when harm to enemy civilians is causally concurrent, or causally posterior, to the securing of a military advantage. In this case, overconfidence in the securing of the military advantage does not push the deliberating agent towards underestimating the relative likelihood of the harm. If an agent overestimates the probability that some event in a causal chain will occur, this does not diminish her ability accurately to assess how comparatively likely it is that concurrent or later links in the chain will occur.

[25]Kahneman and Lovallo, 'Timid Choices and Bold Forecasts', p. 27. See also Don A. Moore and Paul J. Healy, 'The Trouble with Overconfidence', *Psychological Review* 115 (2008), pp. 502–17. Moore and Healy refer to the specific phenomenon that we are interested in as 'overestimation'.

[26]Gilberto Montibeller and Detlof von Winterfeldt, 'Cognitive and Motivational Biases in Decision and Risk Analysis', *Risk Analysis* 35 (2015), pp. 1230–51, at 1232.

[27]Montibeller and von Winterfeldt, 'Cognitive and Motivational Biases', p. 1232. See also Kahneman and Lovallo, 'Timid Choices and Bold Forecasts'.

[28]Dominic D. P. Johnson, *Overconfidence and War: The Havoc and Glory of Positive Illusions* (Cambridge, 2004).

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

In light of these considerations, we propose the following action-guiding principle for ordinary[29] agents who are acting under circumstances of normal uncertainty, and who are considering whether to pursue some military advantage:

Action-guiding principle for ordinary agents in war acting under normal uncertainty (Welfarist Immunity):

- *WI(a).* If an opportunity to secure a military advantage threatens harm to civilians that is causally less removed than the securing of the advantage (viz. the harm to civilians is the means for achieving the military advantage or causally very close to such a means), refrain from pursuing the opportunity.
- *WI(b).* If an opportunity threatens harm to civilians that is causally concurrent with, or posterior to, the securing of the military advantage, calculate the opportunity's expected value. From the set of options that you must not refrain from pursuing, always perform an action with maximal expected value.

This principle, which we refer to as *Welfarist Immunity*, maps on to the distinction that Immunity draws. It rules out inflicting the type of harm that Immunity forbids (WI (a)), and it is permissive with respect to the type of harm that Immunity does not categorically rule out (WI(b)). Unlike Immunity, it yields a clear verdict about the permissibility of harming civilians by riding roughshod over them: it cautions against such harm by grouping it together with harm that is inflicted as a means to a greater good.

Note, though, that Immunity and Welfarist Immunity are not the same kind of principle. Unlike Immunity, which states that targeting civilians is legally and morally impermissible, Welfarist Immunity merely advises well-intentioned agents not to do many of the things that Immunity claims are legally and morally prohibited. Note also that compared to Immunity, Welfarist Immunity is more limited in its applicability – it does not aspire to apply to all moral agents at all times. Only ordinary deliberators, whose irrationality is predictable due to documented biases, are subject to Welfarist Immunity. It follows that welfarism does not comprehensively validate Immunity. Rather, Welfarist Immunity supports many commands that follow from Immunity.

Moreover, Welfarist Immunity's clause WI(b) accords with Immunity's rules on incidentally harming civilians only under a rather permissive reading of Immunity. Clause WI(b) advises agents to pursue whatever alternative they expect will *maximize the general welfare*; Immunity prohibits the infliction of incidental harm that would be *excessive* in relation to the military advantage sought, and requires that agents *minimize* whatever harm they inflict. While WI(b) deems harm excessive if it is not offset by countervailing expected welfare gains, and advises agents against the infliction of unnecessary harm that is not expected to secure any additional welfare benefits, non-welfarist Immunity is frequently interpreted in a more restrictive way.[30] For example, it is commonly argued that incidental civilian harm should be considered disproportionate unless the expected good effects of the military advantage sought outweigh the disvalue of the harm *by a sufficiently wide margin*.[31]

It is not our goal here to settle what is the most appropriate interpretation of non-welfarist Immunity, whether from a legal or a moral perspective. But we note that from

[29]We use the term 'ordinary' to refer to less than fully rational agents who are prone to biases under normal uncertainty.

[30]We thank an anonymous referee for pointing out this fact.

[31]See e.g. Adil Ahmad Haque, *Law and Morality at War* (Oxford, 2017), p. 10.

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

a welfarist perspective, a more restrictive rendering of WI(b) might well be appropriate. More precisely, in section III.2 above, we have suggested that the securing of a military advantage is reasonably conceptualized as the reaching of a *milestone* in the winning of the war. If agents suffer from overconfidence, it is likely that they not only overestimate the probability with which they will secure an advantage, but also the role that the advantage, if secured, would play in the overall war effort. If this is correct, then a more restrictive rendering of WI(b) will be appropriate, namely one that asks agents to *discount* the expected welfare benefits of some military advantage compared to the disvalue of the expected harm to civilians.[32] While determining what discount factor is appropriate depends on our best empirical estimates of agents' overconfidence in war, it is possible that an adequately cautious rendering of WI(b) is very similar in its implications to the rather restrictive interpretations of non-welfarist Immunity that many theorists endorse.[33]

As an action-guiding principle, Welfarist Immunity has moral force if abiding by the principle enables a normally uncertain agent to promote the greater good more successfully than would abiding by any other principle.[34] With respect to the more permissive part of Welfarist Immunity, i.e. clause WI(b), it depends on our best empirical estimates, in the way just outlined, of what precise rendering of the clause has such action-guiding moral force. With respect to the more prohibitive part of Welfarist Immunity, i.e. clause WI(a), we explain in the next section why we believe it has the potential to have action-guiding moral force.

IV.3. Welfarist Immunity as a pre-emptive rule

Clause WI(a) of Welfarist Immunity is based on the idea that under circumstances of normal uncertainty, we might best be able to maximize expected welfare if we abide by a simple, localized decision rule instead of deliberating on the merits of the case. What drives this idea is the insight that deliberating on the merits of the case can sometimes lead us astray in predictable ways.

To get a better grip on this idea, note that the expected welfare-promoting effects of an action that is performed at time t_{act} can be estimated at different points in time. The feeling that, to be properly action-guiding, an estimate should be made as close as possible to t_{act} has to do with the fact that, typically, the epistemic access to the consequences of an action is better the closer in time the agent is to them. But this isn't always or necessarily the case. In general, an agent should thus evaluate the expected value of her actions at a time $t_{optimal}$, i.e. at a time when her epistemic access to the consequences of her actions is as good as she can expect it to get. While typically, $t_{act} = t_{optimal}$, if $t_{optimal}$ occurs before t_{act}, the agent ought to take a future-directed decision at $t_{optimal}$ that helps guide her decision-making at t_{act}. In Luca Ferrero's words, such future-directed decisions are 'tools for the non-manipulative, intrapersonal division of deliberative labor over time. A future-directed decision to ϕ gives rise to a defeasible exclusionary reason to ϕ. This reason is grounded on the default authority that is ideally granted to one's prior self as an "expert" deliberator.'[35]

[32]Cf. Shaw, *Utilitarianism and the Ethics of War*, pp. 138–9.

[33]We thank an anonymous reviewer for suggesting an argument along these lines.

[34]Cf. Jennifer M. Morton, 'Toward an Ecological Theory of the Norms of Practical Deliberation', *European Journal of Philosophy*, 19 (2010), pp. 561–84, esp. at 580.

[35]Luca Ferrero, 'Decisions, Diachronic Autonomy and the Division of Deliberative Labor', *Philosopher's Imprint* 10 (2010), pp. 1–23, at 7.

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

To illustrate, suppose that a stockbroker buys a share just before its price falls by 4%. Further suppose that assets whose price falls so sharply are likely to fall much further before they eventually recover, so that it typically maximizes a stockbroker's gains to sell shares that have fallen by 4% or more within a single day. In such a case, it is important to decide how to react to a shock before the shock occurs. This is so because of a bias called *loss aversion* that makes it psychologically difficult to sell shares whose price has fallen; it is much more agreeable to sell stocks whose price has risen.[36] A stockbroker may therefore do well to impose on herself a rule that commands selling shares whose price has fallen by 4% or more in one day.

Admittedly, such a rule is coarse-grained. After all, shares whose price fell by 4% do not *always* continue falling. At least in principle, the broker might be able to identify whether she is dealing with an exception to the rule once she starts to deliberate on the merits of a particular case. In this sense, abiding by the rule is only a defeasible default strategy: if there is overwhelming evidence that abiding by the rule would lead the agent astray, the agent is not bound to abide by it.[37] Still, once the broker considers whether to hold on to shares whose price has fallen sharply, she should know that self-deception is common and that, despite appearances to the contrary, she probably will not be able to make an improved decision by reconsidering her prior decision to abide by a simple rule. Settling for a rule at $t_{optimal}$ will therefore tend to maximize the broker's expected gains at $t_{optimal}$, and abiding by the rule instead of reconsidering its value at t_{act} will tend to maximize her expected gains at t_{act}.

Based on these considerations, we suggest the following general welfarist action-guiding principle for ordinary agents:

> *General action-guiding principle for ordinary agents*: If you have reason to believe that $t_{optimal}$ occurs at an identifiable time before t_{act}, then at $t_{optimal}$, settle for a decision rule to be employed at t_{act} that maximizes, from the point of view of $t_{optimal}$, the expected value of the action that you will choose to perform at t_{act}. The expected value of an action is the sum of the probability-weighted values of the outcomes that it might bring about. At t_{act}, follow the decision rule. If you do not have reason to believe that $t_{optimal}$ occurs at an identifiable time before t_{act}, then treat t_{act} as $t_{optimal}$.

We suggest, in other words, a welfarist principle according to which performing an action at t_{act} is advisable if choosing to perform the action is in conformity with a rule that was assessed to be welfare-promoting at $t_{optimal}$. In cases where $t_{act} \neq t_{optimal}$, the agent is well advised to abide by the rule at t_{act} because from her best available epistemic standpoint, she expects that deviating from the rule would be counterproductive. It follows that when $t_{act} \neq t_{optimal}$, the proposed general action-guiding principle for ordinary agents is, in a clear sense, Razian: at t_{act}, an ordinary deliberator is subject to the authority of a rule that *pre-empts* the reasons that a better situated or more fully rational deliberator ought to be responsive to at that time.[38]

[36] <http://scienceblogs.com/cortex/2008/09/30/loss-aversion-and-the-stock-ma/>.

[37] See John Horty, 'Reasons as Defaults', *Philosophers Imprint* 7 (2007), pp. 1–28. See also Michael Walzer, *Just and Unjust Wars*, 3rd edn. (New York, 2000), pp. 231–2, where Walzer introduces the notion of a 'supreme emergency' during which the laws of war are suspended.

[38] The notion of pre-emption is seminally developed in Joseph Raz, *Practical Reasons and Norms* (Princeton, 1990), and Joseph Raz, *The Morality of Freedom* (Oxford and New York, 1986).

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

The action-guiding principle suggested here is neither rule consequentialist, nor indirectly consequentialist in some other manner. Instead, it is squarely act consequentialist: its point is to help ordinary agents pick alternatives that they have reason to expect will tend to maximize welfare.[39]

Clause WI(a) of Welfarist Immunity is an example of a decision rule that an ordinary agent should abide by if she finds herself under circumstances of normal uncertainty. The basic thought behind the principle is that under circumstances of normal uncertainty in war, an ordinary agent's overconfidence will lead her to overestimate her chances of securing a military advantage, which in turn will lead her to underestimate probability gaps between the occurrence of causally antecedent harm and the securing of the advantage. Because of this – so the thought goes – the agent does best from a welfarist perspective if she refrains from pursuing military advantages when such gaps are present.

Whether this thought is correct depends on difficult empirical conjectures. We offer the following. From a welfarist perspective, ordinary agents should abide by Welfarist Immunity only if the expected value of doing so exceeds the expected value of always deliberating on the merits of the case. If the agent always deliberates, she will sometimes pursue a military advantage that involves causally antecedent harm when she is in fact justified in doing so, and sometimes she will pursue such an advantage when she is not. We conjecture that on balance, ordinary agents in war are likely to do more harm than good if they consider each option to pursue some military advantage instead of abiding by Welfarist Immunity. Here is why: there are innumerably many ways in which harming civilians *might eventually* lead to the securing of a military advantage *somewhere down the causal chain*. It is easy to tell such stories, e.g. by arguing that the use of human shields has the potential to deter the enemy, or by arguing that terror attacks might undermine civilian morale, and might therefore serve to weaken the enemy's war effort over time. Compared to instances where harming civilians *might eventually* lead to the securing of a military advantage, instances where causally antecedent harm is *closely tied* to the securing of a subsequent military advantage appear few and far between. Hence, the welfare-promoting chances that an agent passes up by following Welfarist Immunity appear to be few, whereas the danger of pursuing misidentified chances – a danger that the agent avoids by following Welfarist Immunity – appears to be pervasive.[40]

[39]See Morton, 'Toward an Ecological Theory of the Norms of Practical Deliberation', p. 578.

[40]There are a fair number of empirical studies that investigate the question of whether targeting civilians has helped further military objectives when civilians have been targeted in the past. If targeting civilians has frequently been ineffective, then this lends support to the idea that in cases where Immunity acts as an inhibitor, it frequently blocks largely ineffective harm whose infliction would not be justified from a welfarist perspective. There is, however, no agreement in the empirical literature on this issue. Influential studies arguing for the ineffectiveness of targeting civilians include Robert Pape, *Bombing to Win: Air Power and Coercion in War* (London, 1996) and Max Abrahms, 'Why Terrorism Does Not Work', *International Security* 31 (2006), pp. 42–78. Influential studies arguing that targeting civilians has sometimes proved effective include Alexander Downes, *Targeting Civilians in War* (Ithaca and London, 2008) and Jason Lyall, 'Does Indiscriminate Violence Incite Insurgent Attacks? Evidence from Chechnya', *Journal of Conflict Resolution* 53 (2009), pp. 331–62. For a careful review of the relevant literature and the methodological issues that empirical studies on the effectiveness of targeting civilians face, see Seth Lazar, 'Necessity and Non-combatant Immunity', *Review of International Studies* 40 (2014), pp. 53–76.

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

V. Conclusion

In this article, we have offered a contingent defence of the strictures of Immunity from a welfarist perspective. Providing such a defence is not a straightforward task, and we grant that our arguments depend on empirical assumptions that may be questioned. Some may feel that this renders our welfarist defence of the strictures of Immunity too limited to be of any real value. After all, if an agent finds herself under ideal deliberative circumstances, this agent is – according to the argument that we have presented – not subject to Welfarist Immunity. This may strike many as deeply implausible, simply because inflicting harm in a case such as *Terror Bombing* appears morally wrong even if we assume that an agent knows what the consequences of her action will be. Moreover, or so it might be argued, new technologies are changing the face of warfare, and as the soberly calculating 'killing by remote control' becomes the new ideal, Welfarist Immunity will prove increasingly obsolete.

In response to these projected criticisms, we offer two thoughts. First, and most importantly, we provide Welfarist Immunity neither in an attempt to replace Immunity, nor in an attempt to provide a substitute for a DDE-based justification of Immunity. We merely note that it is encouraging to see that at least under certain conditions, a welfarist defence of the commands that follow from Immunity is available. Similarly, it is encouraging to see that under certain conditions, welfarism validates most people's intuitive judgement that inflicting riding roughshod harm is harder to justify than inflicting other merely foreseen harm. If, under deliberative uncertainty, welfarism picks up on many of the non-welfarist distinctions that deontologists pay attention to, then this substantially weakens the charges of irrationality and intuition-mongering that welfarists sometimes accuse defenders of deontological principles of.[41] It instead strengthens the hypothesis that different approaches in moral philosophy are climbing the same mountain from different sides, to use Derek Parfit's metaphor.[42]

Second, we accept that from a *welfarist* perspective, an adjustment of the laws of armed conflict could in principle be called for if the face of warfare keeps changing. If it is true that thanks to new technologies, we are slowly moving away from normal uncertainty and towards conditions of ideal uncertainty, then it strikes us as correct that Welfarist Immunity is becoming increasingly obsolete. It seems to us that the big question mark here is with the antecedent: it seems doubtful, to say the least, that humans engaged in armed conflict will be able to rid themselves of deliberative biases anytime soon. As Gilberto Montibeller and Detlof von Winterfeldt point out, overconfidence is pervasive whenever *complex quantitative estimates* need to be made.[43] There is solid evidence that the bias is widespread in the context of defence, financial, legal, and engineering decisions.[44] In other words, one of the main drivers of overconfidence seems to be *incomplete empirical and evaluative information*, and it is doubtful that new technologies will help eliminate these sources of bias in the foreseeable future.

In sum, Welfarist Immunity provides a welfarist defence of the strictures of Immunity under certain imperfect deliberative conditions. Because imperfect deliberative conditions are plausibly pervasive in war, this defence might be more robust than it

[41]See e.g. Joshua D. Greene, 'The Secret Joke of Kant's Soul', *Moral Psychology*, ed. W. Sinnott-Armstrong, vol. 3 (Cambridge, 2008), pp. 35–79.

[42]See Derek Parfit, *On What Matters: Volume I* (Oxford, 2011).

[43]Montibeller and von Winterfeldt, 'Cognitive and Motivational Biases', p. 1232.

[44]Montibeller and von Winterfeldt, 'Cognitive and Motivational Biases', p. 1232. See also Johnson, *Overconfidence and War*.

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press

at first appears. Combined with the fact that Welfarist Immunity provides a similar defence of an injunction against inflicting roughshod harm, this strongly suggests that welfarism is not antithetical to the distinctions that many deontologists take seriously. If we accept that normal uncertainty and the cognitive biases that accompany it are facts of life, then welfarism isn't fundamentally at odds with the idea that some ways of inflicting harm are morally harder to justify than others.[45]

[45]For helpful discussions and comments on earlier versions of this article, we thank Tantum Collins, Matt Hitchens, Ariel Porat, Ofer Malcai, Andreas Schmidt, Arthur Ripstein, Eduardo Rivera-López, Bastian Steuwer, Alex Voorhoeve, audiences at an LSE Choice Group talk in March 2017, at the conference 'Contemporary Just War Theory' at Universidad Torcuato Di Tella' School of Law, and at the 7th Annual Conference of the Oxford Institute for Ethics, Law and Armed Conflict in September 2016, as well as two anonymous reviewers of this journal. Benbaji's research was supported by the Israel Science Foundation, grant number 304/15.

Cite this article: Benbaji Y, Burri S (2020). Civilian Immunity Without the Doctrine of Double Effect. *Utilitas* **32**, 50–69. https://doi.org/10.1017/S0953820819000335

https://doi.org/10.1017/S0953820819000335 Published online by Cambridge University Press