

# **So Close, Yet So Far: Why Solutions to the Closeness Problem for the Doctrine of Double Effect Fall Short<sup>1</sup>**

Dana Kay Nelkin (University of California, San Diego)

Samuel C. Rickless (University of California, San Diego)

## **Abstract**

According to the classical Doctrine of Double Effect (DDE), there is a morally significant difference between intending harm and merely foreseeing harm. Versions of DDE have been defended in a variety of creative ways, but there is one difficulty, the so-called “closeness problem”, that continues to bedevil all of them. The problem is that an agent’s intention can always be identified in such a fine-grained way as to eliminate an intention to harm from almost any situation, including those that have been taken to be paradigmatic instances in which DDE applies to intended harm. In this paper, we consider and reject a number of recent attempts to solve the closeness problem. We argue that the failure of these proposals strongly suggests that the closeness problem is intractable, and that the distinction between intending harm and merely foreseeing harm is not morally significant. Further, we argue that there may be a deeper reason why such attempts must fail: the rationale that makes the best fit with DDE, namely, an imperative not to aim at evil, is itself irredeemably flawed. While we believe that these observations should lead us to abandon further attempts to solve the closeness problem for DDE, we also conclude by showing how a related principle that is supported by a distinct rationale and avoids facing the closeness problem altogether nevertheless shares with DDE its most important features, including an

intuitive explanation of a number of cases and a commitment to the relevance of intentions.

## **1. Introduction**

According to the classical Doctrine of Double Effect (DDE), there is a morally significant difference between intending harm (as a means or as an end) and merely foreseeing (the same amount of) harm.<sup>2</sup> The absolutist version of DDE says that it is never morally permissible to intend harm (even as a means to a good end), though it is sometimes permissible to act in such a way as to foresee (the same amount of) harm.<sup>3</sup> By contrast, non-absolutist versions of DDE hold that it is more difficult to justify intending harm than it is to justify merely foreseeing harm, other things being equal.<sup>4</sup> Although DDE was created and developed by Christian (predominantly Catholic) theologians and philosophers, non-absolutist versions of DDE have become more popular within secular moral philosophy.<sup>5</sup> But there is one difficulty, the so-called “closeness problem”, that continues to bedevil all versions of DDE.<sup>6</sup> The problem is that an agent’s intention can always be identified in such a fine-grained way as to eliminate an intention to harm from almost any situation, including those that have been taken to be paradigmatic instances in which DDE applies to intended harm. For some moral theorists attracted to DDE, this problem is sufficient to justify abandoning DDE as classically understood and replacing it with an intention-based doctrine similar in many ways to DDE that avoids the closeness problem and does a better job of explaining our moral intuitions about cases more generally.<sup>7</sup> However, others think that the closeness problem can be overcome with a better understanding of the nature of intention or the nature of badness.

In this paper, we consider and reject a number of recent attempts to solve the closeness problem. We argue that the failure of these proposals strongly suggests that the closeness problem is intractable, and that the distinction between intending harm and merely foreseeing harm is not morally significant. Further, we argue that there may be a deeper reason why such attempts must fail. The reason is that the rationale that makes the best fit with DDE, namely, an imperative not to aim at evil, is itself irredeemably flawed. While we believe that these observations should lead us to abandon further attempts to solve the closeness problem for DDE, we also argue that this need not lead us to abandon the commitment of DDE to the relevance of intentions for moral permissibility. Thus, we conclude by showing how a related principle that is supported by a distinct rationale and avoids facing the closeness problem altogether nevertheless shares with DDE an intuitive explanation of a number of cases and a commitment to the relevance of intentions.

The plan of the paper is as follows. In section 2, we describe four pairs of cases that are commonly thought to support the classical version of DDE. In section 3, we explain the closeness problem and why it threatens DDE. In sections 4-8, we consider an intriguing variety of solutions to the problem. In section 4, after summarizing some criticisms of simple attempts to avoid the closeness problem, we consider and reject two proposed solutions to the problem: the “loose” solution defended by Bennett (1995) and the “vague” solution defended by Delaney (2008). In section 5, we criticize two different defenses of the thesis that the closeness problem is not really any sort of problem at all, one belonging to Cavanaugh (2006), the other belonging to Masek (2010). In section 6, we criticize a different solution defended by Fitzpatrick (2006), according to which the

proper solution to the closeness problem depends on facts about relations of natural constitution that obtain between different states of affairs. In section 7, we argue against the “thick plans” solution proposed by Hills (2006), one that exploits the claim that the plans of most people usually involve more than a small set of end-directed and means-directed intentions. In section 8, after explaining the difficulties faced by the solution proposed by Hart (1968), we explain what is wrong with a more sophisticated solution (defended by Wedgwood (2011)) that starts from one of Hart’s own premises, namely that there exist true moral generalizations connecting certain types of behavior with some sort of serious injury or death. In section 9, we close the paper by discussing two rationales that are commonly used to defend the moral relevance of DDE, the “aiming-at-evil” rationale and the “using-as-a-means” rationale. We argue that the aiming-at-evil rationale makes good sense of some (though not all) of the main cases commonly used to support DDE, but that the rationale itself is irremediably flawed. We also argue that the using-as-a-means rationale makes even better sense of these cases but supports an alternative principle to DDE (or a non-classical version of DDE) that is a variant of a principle defended by Quinn (1989).

## **2. Cases**

The central claim of DDE, that there is a morally significant difference between intending harm and merely foreseeing harm, is designed to capture and explain our intuitions (i.e., our careful, unbiased judgments) about particular actual and hypothetical pairs of cases. Perhaps the most widely cited pair is (Strategic Bomber, Terror Bomber):

### **Strategic Bomber**

A bomber (fighting a just war) drops a bomb on an enemy munitions factory, intending to destroy the factory and thereby damage the enemy's fighting ability, foreseeing that the fallout from the resulting explosion will cause the death of a number of innocent civilians living near the factory, but not intending these deaths.

### **Terror Bomber**

A bomber (fighting a just war) drops a bomb on an enemy munitions factory, intending the resulting explosion-fallout-caused deaths of a number of innocent civilians living near the factory, as a means of terrorizing the rest of the enemy population into giving up the war effort.

Our intuitions about these cases suggest that what the agent does in Terror Bomber is, other things being equal, more difficult to justify than what the agent in Strategic Bomber does, even though both bombers cause, and foresee that their actions will cause, the same number of enemy deaths. Not unreasonably, proponents of DDE see this pair of cases as evidence that the moral distinction between terror bombing and strategic bombing hinges on the distinction between intending harm and merely foreseeing harm: for while the agent in Terror Bomber intends the death of enemy civilians, the agent in Strategic Bomber merely foresees, without intending, their deaths.

The agents in Terror Bomber and Strategic Bomber both cause harm to innocents by dropping bombs on them. This might suggest to some that the intention/foresight

distinction (and hence, DDE) applies only to pairs of cases involving actions, rather than omissions. But such a suggestion would be mistaken. For there are also pairs of omissions that suggest the existence of a morally significant difference between intending harm and merely foreseeing harm. One such pair is (Direction of Resources, Guinea Pig):

### **Direction of Resources**

There is a shortage of resources for the investigation and proper treatment of a new, life-threatening disease. Doctors decide to cope by selectively treating only those who can be cured most easily, leaving the more stubborn cases untreated. This way, the doctors expect to do a significant amount of long-term medical good.

### **Guinea Pig**

There is a shortage of resources for the investigation and proper treatment of a new, life-threatening disease. Doctors decide on a crash experimental program in which they deliberately leave the stubborn cases untreated in order to learn more about the nature of the disease. By this strategy, they reasonably expect to do as much long-term medical good as they would in **Direction of Resources**.<sup>8</sup>

Our intuitions about these cases suggest that the doctors' omission to treat in Guinea Pig is, other things being equal, more difficult to justify than the doctors' omission to treat in Direction of Resources. And proponents of DDE see the reason for this in the fact that

whereas the doctors in *Direction of Resources* merely foresee the harm suffered in the stubborn cases, the doctors in *Guinea Pig* intend that very harm by intending that the stubborn cases remain untreated. Assuming that the analysis of these cases is correct, it follows that the intention/foresight distinction does not map neatly onto the doing/allowing distinction: sometimes there is an intention to harm via action, but sometimes there is an intention to harm via inaction (or omission). For many (perhaps most) proponents of DDE, DDE is therefore to be distinguished from a different non-consequentialist principle, the Doctrine of Doing and Allowing (DDA), according to which it is more difficult to justify doing harm than it is to justify allowing harm, other things being equal.

Because DDE is meant to be a very general principle, it is essential to test it against not just one or two cases, and it will be helpful in what follows to consider two more pairs of cases commonly thought to support it: (*Trolley*, *Large Man*) and (*Hysterectomy*, *Craniotomy*). In ***Trolley***, a runaway trolley is hurtling down a main track towards five innocent people who are trapped on the main track. A bystander happening by sees that she can save the five by pulling a lever and thereby shunting the trolley onto a side track, where one innocent person happens to be trapped. The bystander pulls the lever as a means of saving the five, foreseeing but not intending the death of the one on the side track. In ***Large Man***, a runaway trolley is hurtling down a main track towards five innocent people who are trapped on the main track. A bystander happening by sees that she can stop the trolley by pushing a large man standing beside her into its path. The bystander pushes the large man into the path of the trolley as a means of saving the five, intending the impact that leads to the large man's death. Some think that DDE derives

some support from (Trolley, Large Man), because (i) other things being equal, it seems more difficult to justify pushing the one onto the trolley tracks in Large Man than it is to justify shunting the trolley onto the side track in Trolley, and (ii) whereas the bystander in Trolley foresees the one's being harmed without intending that harm as a means or as an end, the bystander in Large Man intends the harmful impact that is used as a means to save the five.

In **Hysterectomy**, a doctor removes a pregnant woman's cancerous uterus in order to save her life, foreseeing that the operation will lead to the death of the fetus. In **Craniotomy**, a doctor removes a fetus from a pregnant woman's body in order to save her life by crushing its skull and sucking the remains through the mother's birth canal. Some think that DDE derives support from this pair of cases too, because (i) other things being equal, it seems more difficult to justify the craniotomy than it is to justify the hysterectomy, and (ii) whereas the doctor in Hysterectomy foresees that the fetus will be harmed (indeed, that the fetus will die), the doctor in Craniotomy intends the harmful skull-crushing that is used as a means to save the mother's life.<sup>9</sup>

### **3. The Closeness Problem**

Even assuming that these case-pairs provide intuitive support for DDE, there is one problem for DDE that appears intractable: the closeness problem. The closeness problem arises from the fact that there are cases very similar to, and apparently morally indistinguishable in the relevant respects from, cases such as Terror Bomber in which the relevant agents arguably do *not* intend death or any other kind of harm, but rather intend some other state of affairs that is causally responsible for (or otherwise connected by



some relation other than identity to) death or harm. In Terror Bomber, the bomber is described as intending the deaths of innocent enemy civilians (as means to ending the war by terrorizing the enemy population). But consider now Sophisticated Terror Bomber:

### **Sophisticated Terror Bomber**

A bomber (fighting a just war) drops a bomb on an enemy munitions factory, intending “only that [the enemy civilians’] bodies should be in a state that would cause a general belief that they were dead, this lasting long enough to shorten the war: nothing in that scheme requires that the dismaying condition of the bodies be permanent; so nothing in it requires that the [enemy civilians] become downright dead rather than merely seemingly dead for a year or two.”<sup>10</sup>

The reaction of many is that the actions of the bomber in Sophisticated Terror Bomber are just as difficult to justify, other things being equal, as the actions of the bomber in Terror Bomber, and that it is therefore more difficult, other things being equal, to justify the actions of the bomber in Sophisticated Terror Bomber than it is to justify the actions of the bomber in Strategic Bomber. The problem, though, is that it appears from the description that the bomber in Sophisticated Terror Bomber does not actually intend the *deaths* of the enemy civilians that are impacted by the bombs; nor does it appear that the bomber intends that the enemy civilians who are impacted by the bombs be *harmed* by the impact. The bomber in the case intends that the enemy civilians be sufficiently impacted by the bombs so as to *appear* to be dead *long enough to significantly increase the chances of enemy surrender*. And this appears to be compatible with the bomber’s

not intending that the enemy civilians suffer death or serious harm, whether the bomber believes or does not believe that the impact of the bombs on the enemy civilians' bodies will cause death or serious injury. The structure of the bomber's intentions in Sophisticated Terror Bomber therefore seems indistinguishable in all relevant respects from the structure of the bomber's intentions in Strategic Bomber: both bombers intend a state of affairs (the bombs' exploding in the munitions factory, the bombs' impacting the enemy civilians' bodies in such a way as to make them appear dead) that is causally sufficient for harm (the death of enemy civilians), foreseeing the harm without intending it. But if this is so, then DDE predicts, contrary to our intuitions, that the bomber's actions in Sophisticated Terror Bomber and in Strategic Bomber are equally justifiable, and that it is, other things being equal, more difficult to justify the bombing in Terror Bomber than it is to justify the bombing in Sophisticated Terror Bomber. If our intuitions about the cases should be trusted, it follows that DDE is false.

The fact that the pair (Strategic Bomber, Terror Bomber) is susceptible to the closeness problem is not an artifact of these particular cases. The same sort of difficulty arises in all the other cases too. Indeed, some of these cases already raise the problem *without being redescribed!* In Guinea Pig, for example, what the doctors intend is that the disease progress sufficiently to allow for the acquisition of additional knowledge that will help prevent the disease in a greater number of future cases. In intending that the disease progress in this way, the doctors do not appear to be intending that the relevant patients suffer harm (or death); rather, they merely foresee that the progression of the disease will harm the patients. In this way, the structure of the intentions belonging to the doctors in Guinea Pig appears relevantly similar to the structure of the intentions

belonging to the doctors in Direction of Resources. In both cases, the doctors foresee, but do not intend, harm. Given these facts, it follows from DDE that there is no morally relevant difference between what the doctors refrain from doing in Guinea Pig and what the doctors refrain from doing in Direction of Resources. And yet our intuitions suggest, as we have seen, that it is, other things being equal, more difficult to justify the doctors' omission to treat the stubborn cases in Guinea Pig than it is to justify the doctors' omission to treat the stubborn cases in Direction of Resources. If this is true, then when the structure of intentions in the cases is properly described and understood, the pair (Direction of Resources, Guinea Pig) counts as evidence *against*, rather than as evidence *for*, DDE.

Matters are no different in the case of (Trolley, Large Man) and (Hysterectomy, Craniotomy). In Large Man, the bystander pushes the large man in the path of the trolley as a means of saving the five, but it appears that nothing about the case requires us to suppose that the bystander must intend that the large man suffer harm (or death). We said above that the bystander intends the harmful impact, but this claim is ambiguous; on one reading, the claim appears true, while on another the claim appears false. To intend a harmful impact might be to intend the impact along with its harmfulness, but it might also be to intend an impact that, as it happens, is also harmful. And on the former reading, it seems false, strictly speaking, to say that the bystander intends the harm that comes to the large man. All that the bystander intends is that the body of the large man stop the trolley from hitting the five. As far as the bystander is concerned, whether the trolley actually harms or kills the large man is nothing to her purpose. So here again the structure of intentions of the bystander in Large Man, properly understood, does not differ in any

relevant respects from the structure of intentions of the bystander in Trolley. In Craniotomy, the doctor crushes the fetus's skull as a means of saving the mother's life, but it appears that nothing about the case requires us to suppose that the doctor intends that the fetus suffer harm (or death). The Craniotomy doctor could just as easily be described as intending that the fetus's head be modified in such a way as to permit egress from the mother's body, and this seems compatible with the doctor's *not* intending the harm (or death) that crushing the fetus's skull will bring about. Here too the structure of intentions of the doctor in Craniotomy appears to differ in no relevant way from the structure of intentions of the doctor in Hysterectomy. In both cases, the doctors are aiming to produce a state of affairs (removal from the womb, crushing of the skull) that happens to be causally, but contingently, related to the harm (i.e., death) it brings about. DDE, then, entails that what the bystander does in Large Man (what the doctor does in Craniotomy) is morally no more difficult to justify, other things being equal, than what the bystander does in Trolley (what the doctor does in Hysterectomy). And yet, on careful reflection, the cases strike us as morally distinguishable.

#### **4. "Loose" and "Vague" Solutions**

Bennett (1995, 205-10) describes four attempts at solving the closeness problem, each one of which he finds wanting. We agree with much of his analysis, but for the purpose of evaluating later proposals, it is important to keep some of Bennett's points in mind. So we begin our investigation of suggested solutions by summarizing some of Bennett's main criticisms.

In order to solve the closeness problem, defenders of the classical DDE must show that in cases such as Sophisticated Terror Bomber and Craniotomy the agent of the case really does intend harm of some sort. On the Event Identity proposal, “the collapse of the [fetus’s] head [in Craniotomy] is its death; these are two descriptions of a single event; so the surgeon cannot intend to bring about ‘one’ without intending to bring about ‘the other’” (1995, 205). Similarly, it might be argued that an enemy’s body being impacted by a bomb in Sophisticated Terror Bomber *is* his death (or his being seriously harmed); that the large man’s body being impacted by the runaway trolley *is* his death (or his being seriously harmed); and that the disease’s progression in a stubborn case *is* the patient’s being seriously harmed. But Bennett argues (and we agree) that events are individuated at least in part by subject (or property) and by time of occurrence: *if* event  $e_1$  is subject  $S_1$ ’s having property  $P_1$  and event  $e_2$  is subject  $S_2$ ’s having property  $P_2$  and  $S_1 \neq S_2$ , *then*  $e_1 \neq e_2$ ; and *if* event  $e_1$  occurs at  $t_1$  and event  $e_2$  occurs at  $t_2$  and  $t_1 \neq t_2$ , then  $e_1 \neq e_2$ . It follows from these event individuation principles that the collapse of the fetus’s head, which has as its subject the fetus’s head, is an event that is numerically distinct from the fetus’s death, which has as its subject the fetus; and the same conclusion follows from the fact that “the collapse of the head and the death of the [fetus]...occur a second apart” (1995, 206). The same point applies, *mutatis mutandis*, to Sophisticated Terror Bomber, Large Man, and Guinea Pig. So the Event Identity attempt to solve the closeness problem fails.

On the Act Identity proposal, the act that is the crushing of the fetus’s head is numerically identical to the act that is the killing of the fetus; from this it is then held to follow that it is impossible to intend the crushing without intending the killing. It follows

that in Craniotomy the doctor can be seen as intending harm (or an act of harming). Similarly for Sophisticated Terror Bomber, Large Man, and Guinea Pig. But as Bennett points out, this attempted solution won't work either. For in Hysterectomy, the doctor intends the removal of the fetus from the mother's body. But this removal leads to the fetus's death, and hence counts as a killing. So on the Act Identity proposal, the doctor in Hysterectomy intends a killing by virtue of the fact that she intends the removal of an unviable fetus. Similarly, in Strategic Bomber the Act Identity proposal "implies that the [bomber's] destruction of the factory is his killing of the civilians, because each is identical with certain movements that he makes up in his aeroplane" (1995, 208). The Act Identity proposal therefore entails that the intention/foresight distinction does not enable us to distinguish morally between Craniotomy and Hysterectomy, or between Sophisticated Terror Bomber and Strategic Bomber.

Bennett then discusses two proposals that focus on relations between states of affairs (rather than between events or between acts). Consider the following states of affairs: (i) the fetus's head's being crushed and (ii) the fetus's dying. On the Causal Necessitation proposal, it is acknowledged that these states of affairs are numerically distinct, but, by virtue of the fact that the first causally necessitates the second (in the sense that the first makes the second inevitable), it is impossible to intend the first without intending the second. It follows that the Craniotomy doctor necessarily intends the fetus's dying. The problem with the Causal Necessitation proposal is that a similar relation between the relevant states of affairs holds in Hysterectomy, for "when a hysterectomy is performed early in pregnancy,...it is causally inevitable that the [fetus] will die". Similarly, even though the Causal Necessitation proposal entails that the

bomber in Sophisticated Terror Bomber intends the enemy civilians' dying (by virtue of intending the civilians' being impacted by the bombs), the same proposal also entails that the bomber in Strategic Bomber intends the enemy civilians' dying; for "it is inevitable that if the bombs are dropped the civilians will die" (1995, 209). So the Causal Necessitation proposal, like the Act Identity proposal, fails to distinguish morally between Craniotomy and Hysterectomy, or between Sophisticated Terror Bomber and Strategic Bomber.<sup>11</sup>

Finally, on the Logical Entailment proposal, intending state of affairs S1 is insufficient for intending state of affairs S2 whenever it is logically possible that S1 obtain without S2's obtaining too. On this proposal, it is possible for the doctor in Hysterectomy to intend the removal of the fetus from the mother's body without thereby intending the fetus's dying, for the removal of the fetus does not logically entail its dying. Unfortunately for the Logical Entailment proposal, the same thing is true of Craniotomy, for in that case the doctor's crushing the fetus's skull does not logically entail the fetus's dying either. As Bennett (rightly) points out, "there are worlds where God steps in and restores the ruined head to its former condition, and others where crushing a head is the first step in a helpful curative procedure" (1995, 209-10). Similar remarks apply to Strategic Bomber and Sophisticated Terror Bomber. In Strategic Bomber, the factory's being destroyed does not logically require the civilians' being harmed (or dying); but something similar is true of Sophisticated Terror Bomber, for in that case the civilians' being impacted by the bombs also does not logically require their being harmed (or dying). Here too, then, the Logical Entailment proposal does not enable us to distinguish

morally between Hysterectomy and Craniotomy, or between Strategic Bomber and Sophisticated Terror Bomber.

Bennett himself suggests that the most that could be said in the way of a solution to the closeness problem is this:

The best I can find is rather loose, but it may be the whole truth about our intended/foreseen distinction. Not only is there no chance of turning the ashes back into the building, or the smithereens back into people, or of crushing the baby's head without killing it – these things are what the plain man would call *inconceivable*. We can fairly easily imagine getting technology that would allow bombs to be aimed much more precisely, or would allow a fetus to be brought to term outside the mother's body; whereas the idea of destroying the head but not the baby, or of restoring a person who has been burnt to a cinder, is sheer fantasy. Without denying that it is conceptually possible, something God could do, we have not the faintest idea of what it might be like to have the means to bring it about. That, I suggest, is the 'tight binding' [between states of affairs] we have been looking for. (1995, 213)

It is difficult to know what to make of this suggestion. Bennett recognizes that neither Logical Entailment nor Causal Necessitation will do, and so opts for something in between, a kind of "plain man's" necessitation or entailment that involves not being able to conceive of the means of preventing one state of affairs from following upon another. Delaney (2008, 351) points out, we believe rightly, that Bennett's test is "rather elusive"



and gives us no “clear sense of how to ‘go on,’ even to [relatively simple cases]”. But inasmuch as the test provides us with *any* guidance at all, it seems irremediably flawed. For what strikes one ordinary person as plainly inconceivable at one time might strike another ordinary person at the same time (or at a later time) as plainly conceivable. When precision guidance technology was in its earliest stages, a slightly more knowledgeable person would have been able to conceive of a means of aiming bombs more precisely, at the very time that an ordinary relatively ignorant person would have been unable to conceive of such a means. Conceivability of the relevant sort, then, seems dependent on the amount of knowledge one has. And yet we do not want to say that whether DDE applies to a particular situation depends on the level of knowledge one imagines a representative “plain man” to possess. Before the idea of an incubator was contemplated by anyone, the thought of bringing a fetus to term outside its mother’s body would have been inconceivable to a “plain man”. *Now*, of course, it is conceivable, but only because of the progress of science and general education. But it seems to us inconsistent with the nature of morality that the moral permissibility or impermissibility of a particular kind of conduct should hinge on the amount of scientific knowledge that a hypothetical “plain man” just happens to have.<sup>12</sup>

Having dismissed Bennett’s “rather loose” solution to the closeness problem for its inability to provide moral guidance in simple cases, Delaney (2008) offers a solution that, ironically enough, seems considerably looser and less useful than Bennett’s. Delaney’s suggestion, in essence, is that the intention to A entails the intention to B whenever A and B are “too close” to permit the distinction between intention and foresight to get a hold. When Delaney asks himself whether this proposal is “hopelessly

vague (rather than just vague)”, he pleads “Aristotle’s dictum about the limits of fruitful analysis” and claims that “to [his] knowledge no one has managed to eliminate (or analyze) the notion of felt closeness in favor of (or into) a useful precise standard” (2008, 344). In response to the reasonable criticism that his proposal is circular (because whether A and B are “too close” to permit the distinction between intention and foresight to get a hold is determined by whether the intention to A entails the intention to B), Delaney admits to the circularity, but claims (without explaining why) that it is “benign” and carries “no hint of viciousness” (2008, 344). Ultimately, Delaney concludes that the concept of tight linkage that is sufficient for the intention to A to entail the intention to B is “vague and analytically irreducible” (2008, 345). And to those who might suggest that this shows no more than that he is whistling in the dark, Delaney claims that “if we do not offer some admittedly rough and ready rationale for blocking extremely fine-grained ascriptions of direct intention by using one of the proposed principles of [tight linkage] DDE will simply issue absurd pronouncements about the moral permissibility of patently unacceptable courses of conduct” (2008, 349). But this comment encapsulates our criticism of Delaney’s proposal to a tee. Delaney’s proposal is essentially a non-proposal, a refusal to analyze the concept of tight linkage that purportedly justifies claims of the form “the intention to A entails the intention to B” except by giving a circular account of it that is sufficiently indeterminate and contentless to render principled and consistent application of it impossible. Without such a principle, as Delaney recognizes, we are forced to conclude that DDE is false.

## 5. “Rejection of the Problem” Solutions

Two recent attempts to solve the closeness problem, Cavanaugh (2006) and Masek (2010), rest on the thesis that, properly understood, there is in fact no closeness problem at all. Cavanaugh (2006) argues that each member of any of the common pairs of cases that are designed to illustrate the closeness problem can in fact be distinguished from the other member of the pair in that one, but not the other, involves an intention to harm (in some way or other). Cavanaugh focuses on three pairs of cases in particular: (Active Voluntary Euthanasia, Terminal Sedation), (Craniotomy, Hysterectomy), and (Sophisticated Terror Bomber, Strategic Bomber). Because we agree with (the relevant parts of) Cavanaugh’s analysis of Active Voluntary Euthanasia (2006, 108-09), we will concentrate here on the other two pairs of cases.

Cavanaugh begins his analysis of the closeness problem by explicating a theory of intention, as distinguished from mere foresight, that is mostly borrowed from Bratman (1987). The account rests on five features that differentiate intention from mere foresight. In the first place, intention is a “volitional commitment to a plan of action” (2006, 107). For example, “in intending to go to Venice, [we have] a plan to go to Venice to which [we] commit [ourselves] and in accordance with which [we] save money, reserve vacation time, arrange transportation, study Italian, and perform myriad other acts” (2006, 93-94). As such a commitment, an intention “constrains [us] (rationally) from having other, incompatible intentions, and, thereby, from acting in ways inconsistent with [our] plan (2006, 94). Mere foresight, by contrast, does not involve a volitional commitment to an action plan, for it is no more than an “awareness of causal relations” (2006, 95). Second, “intention characteristically causes both deliberation [about effective means] and

further intentions; foresight does not” (2006, 107). For example, if we intend to go to Venice, we will ask ourselves “how can [we] bring about this intended end?” (2006, 98), and thus we will deliberate about the best means of getting to Venice. By contrast, merely foreseeing that we will go to Venice does not cause us to deliberate about the best means of getting there. Third, “the intention of a means solves deliberation’s problem [of how to effect an intended end]; foresight does not” (2006, 107). For example, the intention to save money, reserve vacation time, and arrange transportation to Venice solves the problem of how to bring about the intended end of going to Venice. Fourth, “intent and foresight differ in terms of their directions of fit; [a]s practical knowledge, intent defines an act as a success or failure; as speculative knowledge, foresight does not” (2006, 107). For example, our intent to go to Venice defines our getting to Venice as a success and our not getting to Venice as a failure. By contrast, mere foresight that we will go to Venice does not count as success if true or failure if false. And fifth, “if one were to lack a specific intent, one’s act could not be named by that intent and one’s conduct typically would differ from what it would be were that intent present; [h]owever, if one were not to foresee some consequence of one’s act, one’s act would not need to be redescribed nor would one’s behavior typically differ” (2006, 107). Thus, our intention to go to Venice differentiates our act of going to Venice from its merely happening that we end up in Venice, distinguishes the act of getting on a Venice-bound plane as an act of going to Venice, as opposed to an act of, say, irritating our friends by our absence, and informs behavior in that we wouldn’t get on a Venice-bound plane if we didn’t have the relevant intention. By contrast, “foresight neither defines one’s act nor informs one’s conduct” (2006, 105).

Cavanaugh then applies his analysis to Craniotomy and Sophisticated Terror Bomber. In the case of Craniotomy, Cavanaugh argues that the doctor is volitionally committed to the goal of saving the mother's life and to a plan of action that is the outcome of deliberation about the best means of achieving that goal. The plan the doctor decides on, one that solves his deliberative problem, defines her actions as successful or not, and informs her conduct, involves "cut[ting] a hole in the baby's skull, remov[ing] its brain, crush[ing] its skull, and then remov[ing] its corpse from the birth canal" (2006, 112). Cavanaugh concludes, given Bratman's theory of intention, that the doctor "necessarily intends to destroy the baby's head" (2006, 113). Moreover, Cavanaugh assumes that to destroy a baby's head "is to intend grievous bodily harm to the baby", and from these facts it follows that any doctor who performs a craniotomy to save the mother's life intends serious harm to the baby whose head she crushes. In the case of the Sophisticated Terror Bomber, Cavanaugh argues that the bomber is volitionally committed to the goal of terrorizing the enemy population into giving up the war effort, and to a plan of action that involves making enemy civilians appear dead as a means to his end. This plan, which is the outcome of deliberation about how best to terrorize the enemy population, which solves his deliberative problem, defines his actions as successful or not, and informs his conduct, involves "killing the non-combatants" as a means of making them appear dead. (What more effective means could there be?) Cavanaugh concludes that "the [sophisticated] terror bomber requires the civilian deaths as a means to his intended end...He necessarily intends...to kill and grievously harm [the civilians]" (2006, 115).

Cavanaugh's presentation of Craniotomy and Sophisticated Terror Bomber suggests that he treats Bratman's theory of intention as a set of principles from which it is possible to infer that the agents intend to harm or kill as a means to their respective ends. But, in fact, Bratman's theory of intention, taken on its own, does not entail the results for which Cavanaugh argues. What Bratman's theory entails is no more than (i) that the Craniotomy doctor intends to modify the shape of the fetus's head as a means of extracting its body safely from the birth canal, and (ii) that the bomber in Sophisticated Terror Bomber intends to impact civilians with bombs as a means of making them appear dead (long enough to terrorize the enemy leaders into giving up the war effort). By itself, then, Bratman's theory entails neither that the doctor intends to *harm* the fetus nor that the sophisticated terror bomber intends to *harm* the enemy civilians who are impacted by the bombs. To obtain these results, Cavanaugh must make two additional claims, namely that intending to destroy a fetus's head entails intending to harm the fetus, and that intending to impact enemy civilians with bombs entails intending to harm them. And to bolster these claims, Cavanaugh assumes that to destroy a fetus's head *is* to harm the fetus, and that to impact people with bombs *is* to harm them.

The problem with Cavanaugh's proposal is that from the fact that the act of A'ing is identical to the act of B'ing and the fact that X intends to A, it does not follow that X intends to B. If it did, then we would be forced to say that the *Hysterectomy* doctor intends to harm the fetus and that the bomber in *Strategic Bomber* intends to harm the enemy civilians who are impacted by his bombs. For it is clear that the Hysterectomy doctor intends to remove the (unviable) fetus from the mother's body, and that the act of removing such an unviable fetus *is* an act of harming (indeed, killing) it; and it is also

clear that the bomber in Strategic Bomber intends to impact the enemy civilians with bombs, and that the act of impacting such persons with bombs *is* an act of harming (indeed, killing) them. And yet, as Cavanaugh emphasizes, we want to say neither that the Hysterectomy doctor intends to harm (or kill) the fetus nor that the bomber in Strategic Bomber intends to harm (or kill) enemy civilians. This is no more than to emphasize the same point that Bennett makes against the Act Identity proposal for solving the closeness problem (see section 4 above).

Also appealing to an account of the nature of intention, Masek (2010), following Marquis (1991), adopts a 'strict' theory according to which "an effect is intended (or part of the agent's plan) if and only if the agent A has the effect as an end or believes that it is a state of affairs in the causal sequence that will result in A's end" (2010, 569). Applying the strict theory (which is compatible with Bratman's theory), Masek accepts that in cases such as Large Man and Craniotomy, the agent of the case does not intend that the relevant victim, large man or fetus, suffer harm or death.<sup>13</sup> This puts Masek in a potentially precarious theoretical position, however, because it follows from his theory of intention that the structure of intentions in Large Man is relevantly indistinguishable from the structure of intentions in Trolley, and that the structure of intentions in Craniotomy is relevantly indistinguishable from the structure of intentions in Hysterectomy. Just as the bystander who diverts the trolley does not intend the death of the one as a means to saving the five, so the bystander who pushes the large man in the path of the trolley does not intend the large man's death as a means to saving the five; and just as the doctor who removes the unviable fetus from the mother's body does not intend the death of the fetus as a means to saving the mother's life, so the doctor who crushes the fetus's skull in order

to extract the fetus's body from the birth canal does not intend the fetus's death as a means to saving the mother's life. It appears to follow, then, that DDE can get no purchase at all in these pairs of cases.

Masek finesses this problem by biting the bullet. He accepts that the intention/foresight distinction does not, in fact, ground a moral distinction between Large Man and Trolley or a moral distinction between Craniotomy and Hysterectomy. But he holds that there *is* a moral difference between the members of these pairs of cases, a moral difference grounded in a moral principle *unconnected to DDE*. The relevant moral principle is that one should not cause harm unjustly (2010, 571). The problem with Craniotomy, for Masek, is not that the doctor intends to kill or harm the fetus: the problem is simply that the doctor harms the fetus unjustly (by crushing its skull). And the problem with Large Man, for Masek, is not that the bystander intends to kill or harm the large man: the problem is simply that the bystander harms the large man unjustly (by pushing him in the path of the trolley).

Masek claims that the moral principle that one should not cause harm unjustly is not the *only* moral principle that can explain the moral difference between members of case pairs that are otherwise indistinguishable with respect to the structure of intentions. To see this, Masek compares Trolley to another case we might call **Living Patient Transplant**. In Living Patient Transplant, surgeon Lily “anaesthetizes” healthy innocent patient Otto, “removes [his] organs without his consent”, thereby “caus[ing] his death”, and then transplants the organs into five very sick patients, thereby saving their lives (2010, 570). According to Masek's ‘strict’ account of intention, “Lily does not intend Otto's death” (2010, 570), and hence the structure of Lily's intentions does not differ



relevantly from the structure of the bystander's intentions in Trolley. But Masek thinks that Living Patient Transplant is morally distinguishable from Trolley by virtue of the fact that Lily *steals*, whereas the bystander doesn't. The relevant distinguishing moral principle here, then, is that one should not steal.<sup>14</sup>

We agree with Masek that it is not by appeal to DDE that we should distinguish morally between Trolley and Large Man, or between Hysterectomy and Craniotomy. But we do not think that the relevant distinguishing principles are of the form that Masek identifies ("one should not steal", or "one should not cause harm unjustly", or "one should not commit insurance fraud" (2010, 571)). The best way to bring out this point is to focus on Guinea Pig and Direction of Resources. It seems clear to us that it is, other things being equal, more difficult for the doctors to justify withholding treatment in Guinea Pig than it is for them to withhold treatment in Direction of Resources. And yet it seems that no appeal to intention-structure or to the sorts of quotidian moral principles to which Masek appeals can help us explain the moral difference between the cases. According to Masek's 'strict' theory of intention, the doctors in Guinea Pig do not need the existing patients to be harmed (or die) in order to achieve their end of learning more about the disease and curing many more future patients: all they intend is that the disease progress in the stubborn cases. In this way, from the point of view of the 'strict' theory, the doctors in Guinea Pig do not differ (in any relevant way) in their intentions from the doctors in Direction of Resources. For the latter doctors too do not need the 'stubborn' patients to be harmed (or die) in order to achieve their end of treating the 'easy' patients. And given that the doctors in both cases do not cause harm of any kind (whether unjustly or otherwise) and do not lie, cheat, or steal, it is difficult to imagine what sort of run-of-

the-mill moral principle Masek could possibly use to account for the morally relevant difference between the cases. We conclude from this investigation that neither Cavanaugh's nor Masek's proposed solution to the closeness problem succeeds.

## 6. A "Natural Constitution" Solution

Fitzpatrick (2006) offers a metaphysically sophisticated, careful, and nuanced solution to the closeness problem. Like Bennett, Fitzpatrick claims that the proper analysis of closeness requires us to focus on the states of affairs that are potential objects of intentions. The essence of his proposal is this:

[I]f the relation between two states of affairs is known to the agent, natural, and constitutive rather than merely causal, then we cannot properly speak of an agent's intending the one while merely foreseeing but not intending the other. (2006, 603)

So Fitzpatrick offers a sufficient condition for A's intending that S1 occur being sufficient for A's intending that S2 occur, when  $S1 \neq S2$ : (i) S1 constitutes S2, (ii) the constitution relation that relates S1 and S2 is natural, and (iii) A knows (i) and (ii). Each of (i) and (ii) requires explication before Fitzpatrick's proposal can be evaluated.

Let us begin with the constitution relation. In the first place, Fitzpatrick insists that constitution is not identity, logical entailment, or causal necessitation. For example, Fitzpatrick claims that a man's being blown to bits constitutes his being killed (2006, 595). But in saying this, Fitzpatrick does not mean that being blown to bits *is* or *logically*

*entails* being killed, for it is logically possible to be blown to bits without being killed; nor does he mean that being blown to bits causally necessitates (or causes) being killed, for he agrees with Bennett that causal necessitation between events does not solve the closeness problem (2006, 598). This much is clear. But, in the second place, Fitzpatrick does not offer a *theory* of constitution *between states of affairs*. Instead, he offers a significant number of paradigmatic examples to illustrate the kind of relation he has in mind. It is unclear whether Fitzpatrick thinks that the constitution relation is not susceptible of definition (because, say, it is a prototype concept or family resemblance concept) or whether he thinks that the constitution relation is definable but the details of the definition unnecessary or distracting.<sup>15</sup> In any event, Fitzpatrick offers the following (in addition to the relation between being blown to bits and being killed) as examples of constitution: (a) “the fetus’s skull’s being crushed *constitutes* its being seriously injured or killed” (2006, 595); (b) “people’s being impacted with ordinary lethal bombs in such a way as to make them appear dead [constitutes] their being killed” (2006, 591); (c) “an area’s being obliterated [by carpet bombing] is *constitutive* of its occupants’ being destroyed” (2006, 600); (d) a bullet’s going through (the heart of) a human shield constitutes his being killed (2006, 600); and (e) if there is “a convention in place stipulating that a student who receives an ‘F’ in any class is thereby ineligible for financial aid”, then “having an ‘F’ *constitutes* ineligibility for financial aid” (2006, 602). As Fitzpatrick conceives of it, then, the constitution relation between states of affairs is metaphysically both weaker than numerical identity (and logical necessitation) and stronger than causal necessitation.

Next, Fitzpatrick distinguishes between two kinds of constitution relations (or two ways in which constitution relations are determined): natural(ly) and conventional(ly). A constitution relation is conventionally determined (or simply conventional) when it is “determined by merely *conventional* arrangements involving the agency of others” (2006, 602) or “depends on conventions involving other people’s agency” (2006, 603). By contrast, a constitution relation is naturally determined (or simply natural) when it is “simply a matter of *natural* fact” (2006, 602). Applying this account of the natural/conventional distinction, Fitzpatrick claims that (a)-(d), along with the relation obtaining between a man’s being blown to bits and his being killed, are examples of *natural* constitution, while (e) is an example of *conventional* constitution. For the fact that a man’s being blown to bits constitutes his being killed, or the fact that a fetus’s skull’s being crushed constitutes its being seriously injured or killed, is a matter of natural fact; but the fact that a student’s receiving an ‘F’ constitutes her ineligibility for financial aid is the result of university policies imposed and enforced by other agents.

Now, as Fitzpatrick argues, when state of affairs S1 constitutes state of affairs S2, an agent A’s intention that S1 obtain is not sufficient for A’s intention that S2 obtain unless S1 *naturally* constitutes S2 and A knows that S1 naturally constitutes S2. For Fitzpatrick recognizes that, e.g., it is “quite possible...to intend [a particular] student’s receiving an ‘F’...without thereby intending his being rendered ineligible for financial aid” (2006, 602). So when the constitution relation that obtains between S1 and S2 is conventional, A’s intending that S1 obtain need not entail A’s intending that S2 obtain. Moreover, Fitzpatrick recognizes that, even when the constitution relation that obtains between S1 and S2 is *natural*, it is possible for A to intend that S1 obtain without A’s

intending that S2 obtain *if A does not know that S1 naturally constitutes S2*. As he writes: “[T]he fetus’s skull’s being crushed *constitutes* its being seriously injured or killed..., so that it is impossible to aim at the first without thereby aiming at the second (except in the unlikely case of ignorance of the relation between skull crushing and injury or death...)” (2006, 595). The appropriate ‘transfer’ of intentions from S1 to S2 therefore requires that the relevant agent be aware of the fact that S1 (naturally) constitutes S2.

With the help of his natural constitution account, Fitzpatrick can solve the closeness problem in the cases with which we are familiar. In Hysterectomy, the doctor intends the fetus’s being removed from its mother’s body, but this intended state of affairs causes, rather than constitutes, the fetus’s being seriously injured or killed; thus whereas the Craniotomy doctor intends the fetus’s skull’s being crushed and thereby (because of the natural constitution account) intends the fetus’s being seriously injured or killed, the Hysterectomy doctor intends no such thing. In Trolley, the bystander intends the trolley’s being shunted onto the side track, but this intended state of affairs causes, rather than constitutes, the one’s being killed; thus whereas the Large Man bystander intends the large man’s body’s stopping the trolley from crushing the five and thereby (because of the natural constitution account) intends the large man’s being seriously injured or killed, the Trolley bystander intends no such thing. And in Strategic Bomber, the agent intends the bombs’ destroying the enemy factory, but this intended state of affairs causes, rather than constitutes, the enemy civilians’ being seriously injured or killed; thus whereas the agent in Sophisticated Terror Bomber intends the enemy civilians’ being impacted by the bombs in such a way as to appear to be dead and thereby (because

of the natural constitution relation) intends the civilians' being seriously injured or killed, the agent in Strategic Bomber intends no such thing.<sup>16</sup>

Despite its elegance and theoretical advantages, we do not find Fitzpatrick's natural constitution account sufficiently persuasive. In the first place, the concept of constitution is theoretical, not ordinary, and there are questions about whether a series of paradigmatic correct (and incorrect) applications of a theoretical concept can provide us with sufficient elucidation of it. With respect to any ordinary concept C ('bird', 'rock', 'chair', and so on), it is usually clear from the panoply of paradigmatic examples of C's and non-C's how to 'go on' and identify new cases as belonging in the C or non-C category. But theoretical concepts ('prime number', 'omniscient', 'gene') usually require a definition to back the classification of new instances. So Fitzpatrick's reliance on an unexplicated theoretical concept ('constitution', as a relation applying to states of affairs) casts doubt on the integrity and usefulness of his account.

This problem is exacerbated by puzzles generated by the particular paradigms, (a)-(e), that Fitzpatrick uses to elucidate the meaning of 'constitution'. Paradigm (c) is that "an area's being obliterated [by carpet bombing] is *constitutive* of its occupants' being destroyed" (2006, 600). But Fitzpatrick also tells us that a spot on a target *partly* constitutes the target, and because of this it is impossible to "aim at a spot on a target without aiming at the target it partly constitutes" (2006, 595). So, on the one hand, Fitzpatrick claims that a whole's (e.g., an area's) being destroyed constitutes a part's (the area's occupants) being destroyed; and, on the other hand, he claims that a part (e.g., a spot on a target) partly constitutes a whole (the target itself). The latter claim suggests that a part's being destroyed partly constitutes a whole's being destroyed. But this seems

inconsistent with the claim that a whole's being destroyed constitutes a part's being destroyed. Surely it cannot be *both* that X (wholly) constitutes Y *and* Y partly constitutes X.<sup>17</sup>

In the second place, there are issues related to the knowledge condition for its being impossible for A to intend that S1 obtain without A's intending that S2 obtain: this is the claim that A knows that S1 naturally constitutes S2. First, we note that this condition is sufficient *on its own* to generate the relevant principle of intention transfer; for the other two conditions are that S1 constitutes S2 and that the constitution relation between S1 and S2 be natural, and knowledge that *p* entails (that it is true) that *p*. Second, the fact that *knowledge* of a certain kind is sufficient for intention transfer immediately raises the question whether mere *belief (or justified belief) in the absence of knowledge* is sufficient for intention transfer. Suppose, for example, that a doctor *believes*, perhaps with some justification, but also as it happens erroneously, that a fetus's being removed from its mother's body *naturally constitutes* its being killed. Shouldn't the question of whether the doctor's intending the fetus's being removed is sufficient for the doctor's intending the fetus's being killed be as responsive to the doctor's (justified) beliefs as it is responsive to the doctor's knowledge? To see this, consider the following analogy. Suppose that Sally intends the Caped Crusader's being rewarded, and Sally (justifiably, though erroneously) believes that Justin is the Caped Crusader. Shouldn't it follow from this that Sally intends *Justin's* being rewarded? If this is how belief interacts with intention when the numerical identity of states of affairs is at issue, why shouldn't it also be how (justified) belief interacts with intention when the natural constitution of states of affairs is at issue? But if (justified) belief functions in this context in the way that

Fitzpatrick assumes only knowledge does, then it is more accurate to say that it is the fact that A (justifiably) believes (perhaps even erroneously) that S1 naturally constitutes S2 that suffices for its being impossible for an agent to intend that S1 obtain without intending that S2 obtain.

In the third place, Fitzpatrick's claim that there is a morally relevant difference between conventional and natural constitution seems *ad hoc* and insufficiently well defended. Fitzpatrick claims that when (it is known that) S1 conventionally constitutes S2, but not when (it is known that) S1 naturally constitutes S2, it is possible for an agent to intend that S1 obtain without intending that S2 obtain. So, for example, because (it is known that) a student's receiving an 'F' only *conventionally* constitutes the student's being ineligible for financial aid, it is possible for an agent to intend a student's receiving an 'F' without intending the student's being ineligible for financial aid. By contrast, as we have seen, because (it is known that) the fetus's skull's being crushed *naturally* constitutes the fetus's being seriously injured or killed, it is *impossible* for an agent to intend the fetus's skull's being crushed without intending its being seriously injured or killed.

Fitzpatrick recognizes that some might balk at the moral significance of the conventional/natural distinction, and thus sees the need to "give a principled explanation for why [a case would allow] for the intend/foresee distinction despite the constitutive relation between states of affairs" (2006, 602). Fitzpatrick's "principled explanation" is that "the merely *conventional* status of the constitutive relation is what opens up the space for the intend/foresee distinction here, where there would not be such space in the



case of a naturally constitutive relation; and it does so at least partly because of the involvement of others' agency". In the grading case, says Fitzpatrick,

[t]here is a sense in which I can legitimately say: 'I am simply giving you the grade you deserve for my class, and the rest is [the university administration's] doing, which I would prevent if I could.' By contrast, it would be absurd to try to say something parallel in the [craniotomy case], claiming that I am simply...crushing the fetus's skull and the rest is someone else's (the universe's?) doing. There is no other agency in the natural relation cases, so that the agent is the only one to whom the harm is to be attributed. (2006, 603)

But where Fitzpatrick sees absurdity here, we see room for a completely coherent response. What explains the fact that a skull's being crushed constitutes being seriously injured or killed is a set of natural laws. Now some (we included) believe that no agent created the natural laws or structured the universe in accordance with them. But others believe that the natural laws are created by God (or some other powerful being) and that they hold of the universe because God (or some other powerful being) chooses that they should. Such people would straightforwardly deny Fitzpatrick's claim that "[t]here is no other agency in the natural relation cases", for they would say, not that the fact that the crushing of a fetus's skull leads to serious injury or death is the *universe's* doing, but rather that this fact is *God's* doing (or the doing of some other powerful being).

Suppose, then, that Sally thinks that the laws of nature governing skull-crushing and serious injury are created and imposed on the universe by God. According to

Fitzpatrick's "principled explanation" for why conventional constitution "creates conceptual space for the intend/foresee distinction", it follows that Sally, but not an atheist, *can* intend the fetus's skull's being crushed without intending the fetus's being seriously injured or killed. It would then follow that life-saving craniotomies performed by *theist* physicians are easier to justify than life-saving craniotomies performed by *atheist* physicians. This is a morally bizarre result, and one we believe Fitzpatrick would (and should) be unwilling to embrace. But if Fitzpatrick's "principled explanation" fails, then his natural constitution account, at least as applied to the Craniotomy/Hysterectomy pair and others like them, rests on an unprincipled (i.e., *ad hoc*) distinction between conventional and natural constitution. We conclude, from the three reasons we have provided, that Fitzpatrick's natural constitution solution to the closeness problem fails.

### 7. The "Thick Plans" Solution

Unlike Fitzpatrick, Hills (2006) does not believe that there is a straightforwardly metaphysical (or epistemologico-metaphysical) solution to the closeness problem. Borrowing Bratman's 'thin' conception of intention, Hills recognizes (and we agree) that "[i]ntentions are intensional: however 'close' X is to Y, even if X is *identical* with Y, it is possible for an agent to intend that X and not to intend that Y" (2006, 265). Indeed, Hills argues that it is even possible for an agent who *knows that* X is identical to Y to intend that X without intending that Y, as long as the agent has not "drawn the requisite inference" (2006, 266). It follows that, for Hills, there is no analytic or necessary principle of 'intention transfer' that *requires* that an agent who intends that X also intend that Y when X and Y are sufficiently 'close' (and other metaphysical or epistemic

conditions are met). Hills therefore accepts that it is *possible* for the agent in Sophisticated Terror Bomber *not* to intend that the enemy civilians be killed, that it is *possible* for the bystander in Large Man *not* to intend that the large man be killed, and that it is *possible* for the doctor in Craniotomy *not* to intend that the fetus whose skull she crushes be killed. *Some* Sophisticated Terror Bomber cases, then, are going to be morally indistinguishable from Strategic Bomber; *some* Large Man cases are going to be morally indistinguishable from Trolley; and *some* Craniotomy cases are going to be morally indistinguishable from Hysterectomy. But, as Hills argues and emphasizes, *not all*.

Hills claims that most agents adopt ‘thick’, rather than ‘thin’, plans of action, inasmuch as their plans involve many more intentions than merely the intention of achieving an end and the intention of achieving the necessary means to that end (2006, 267). Given that this is true of most agents, Hills reasons that whenever one is faced with a sophisticated terror bomber (or bystander in a Large Man case, or doctor in a Craniotomy case), there will usually be good (though defeasible) reasons to believe that the relevant agent does in actual fact intend innocent people’s being killed. It follows from Hills’ theory that even if *some* Sophisticated Terror Bomber agents do not intend innocent people’s being killed, *most* do; *mutatis mutandis* for the other cases.

What are the reasons Hills sees for thinking that most bombers in the Sophisticated Terror Bomber situation intend enemy civilians’ being killed? The answer is that “killing [civilians] is a very obvious [salient sufficient] means of achieving [the aim of making them seem dead] and we have no evidence that [the sophisticated terror bomber] has adopted an alternative means” (2006, 269). Hills is likely assuming more generally that if M is (known to be) a very obvious [salient sufficient] means of achieving

E and there is no evidence that agent A has adopted a means to achieving E other than M, then it is reasonable to interpret A as having adopted M as a means to achieving E. (This principle, Hills emphasizes, has defeaters, for it is always possible that an agent “not have chosen the most salient means to his end” (2006, fn. 20, 280).) Hills then argues that “[b]y judging what most agents would choose to do in the circumstances, we can interpret the intentions of [the sophisticated terror bomber] in accordance with both common sense and the traditional [account of the case] by defenders of DDE” (2006, 277).

The main problem with Hills’ analysis of Sophisticated Terror Bomber, we believe, is that it does not match our intuitive reactions to the case. Hills’ theory predicts that an ordinary person aware of all the relevant facts *should* react to Sophisticated Terror Bomber, not by saying that the agent’s action is more difficult to justify, other things being equal, than what the bomber does in Strategic Bomber, but that *the evidence, such as it is, suggests that it is more likely than not (or just very likely) that* the agent’s action is more difficult to justify, other things being equal, than what the bomber does in Strategic Bomber. But this is not how ordinary people react to this pair of cases. By and large, ordinary people are quite confident that the bombing in Sophisticated Terror Bomber *is*, other things being equal, more difficult to justify than the bombing in Strategic Bomber. People who are asked to evaluate the agents’ actions in these cases are not epistemically diffident in the way that Hills’ theory suggests that it would be appropriate for them to be. In particular, they do not qualify their evaluations by acknowledging that their judgments about the cases are defeasible or that further evidence about the psychological states of the agents of the cases might well lead them to

issue *opposite* evaluations. If Hills is right, then our intuitive reactions to Sophisticated Terror Bomber and Strategic Bomber do not make sense.

Interestingly, Hills offers a different set of reasons for thinking that most Craniotomy doctors intend the death of the fetuses whose skulls they crush. She begins by noting that most people are not unusually callous or cruel, and that it is unusually callous or cruel to “choose a plan that might involve watching [one’s] victim die slowly and painfully as a result of injuries that [one oneself] had inflicted on him, when [one] could have chosen to kill him instantly” (2006, 272). From this, it follows directly that most people would avoid plans that might involve watching their victim die slowly and painfully as a result of injuries that they themselves had inflicted on him; and if they know that the only way to avoid this sort of plan while still aiming at their ultimate end is to aim at causing and hastening the victim’s death, then they will intend that their victim be killed. As Hills sees it, this is exactly what it is reasonable for us to suppose is true of most Craniotomy cases. Faced with such a case, Hills thinks we reason as follows. Once the Craniotomy doctor adopts the plan of crushing the fetus’s skull to save the mother’s life, she “must intend to harm the foetus, with the foreseen consequence that her victim will either die or be left in a state worse than death; or she must intend to kill.” But given that it is unusually callous or cruel to leave one’s victim in a state worse than death and that most people are not unusually callous or cruel, it follows that the doctor “would choose to ensure that [her] victim died as quickly as possible” (2006, 275).

This reasoning runs through the assumption that most people are not unusually callous or cruel. What this means is that if Hills is right, then our intuitive reactions to Craniotomy *should* be sensitive to evidence regarding the degree of callousness we take

the Craniotomy doctor to possess. But the fact is that our intuitions are not sensitive in this way: just as we do not stop to ask about further information about the psychological states of sophisticated terror bombers as a necessary step towards providing a confident moral assessment of their actions, so we do not stop to ask about the degree of callousness doctors possess before passing confident judgment on their actions. Moreover, Hills' account entails that the *more callous* we take a Craniotomy doctor to be, the *less inclined* we will be to judge her to be intending the fetus's being killed, and that the *less callous* we take such a doctor to be, the *more inclined* we will be to judge her to be intending the fetus's being killed. But this result does not fit our intuitions or the intuitions of ordinary folk who are asked about these cases. Indeed, these consequences of Hills' account appear to be the opposite of what one would expect, for in general we are more inclined to judge X as intending another's being killed the more callous we take X to be, and we are less inclined to judge X as intending another's being killed the less callous we take X to be.

Beyond the difficulties Hills faces in accounting for our intuitions in Sophisticated Terror Bomber and Craniotomy, there is an additional difficulty stemming from the inability of her account to handle cases such as Guinea Pig. Recall that in Guinea Pig, doctors do nothing to treat a disease in the more stubborn cases (even though they could) because they want to learn from the progression of the disease how it might best be cured or its effects much better palliated for a much larger number of future patients. As we argued in section 3, when asked whether they intend the serious injuries or deaths of their patients, these doctors might well say, in all sincerity, that they do not. They might say, rather, that they intend to prevent the disease from injuring or killing a

larger number of patients in the future, and that their intended means of doing so is to learn more about the disease by watching it progress in the stubborn cases: that the patients are seriously injured or harmed by the disease is therefore not to the doctors' purpose.

What can Hills say about Guinea Pig? Hills' general strategy is to assume that most people operate with 'thick', rather than 'thin', plans of action, and that such plans incorporate many more intentions than merely the intention to achieve a particular end and the intention to use some necessary means to do so. In *Sophisticated Terror Bomber*, Hills posits that most bombers in such a position would intend the most salient sufficient means for making enemy civilians appear dead, namely that the civilians actually be killed. In *Craniotomy*, Hills posits that most doctors in such a position would intend to avoid leaving the fetus to a fate worse than death, and would therefore intend its being killed as quickly as possible. But these suggestions do not transfer over to *Guinea Pig*. In *Guinea Pig*, the most salient sufficient means of achieving the end of preventing the disease from ravaging future patients is the acquisition of information about the biochemical pathways that define the disease, information that could help researchers discover ways of blocking or curing the disease. So the "most salient sufficient means" gambit does not provide us with any reason for thinking that most *Guinea Pig* doctors intend that their 'stubborn' patients die. And because the doctors in *Guinea Pig* are not doing anything to their patients that might lead them to a fate worse than death, the assumption that most doctors are not callous entails at most that most doctors in *Guinea Pig* situations would hope that their 'stubborn' patients neither suffer nor die. Not only is this *hope* not an *intention*, its content is the exact opposite of patients' dying (whether

more quickly or less quickly). For all that Hills offers us, then, there is no reason to suppose that most Guinea Pig doctors' 'thick' plans entail a commitment to their 'stubborn' patients being seriously injured or killed. Given that we cannot think of any other reasons that would justify the attribution of the relevant sort of disfavored intention to the Guinea Pig doctors, the result is that Hills' theory offers us no way of distinguishing in practice between Guinea Pig and Direction of Resources. And yet the intuition that it is more difficult to justify the doctors' omission to treat in the former than it is to justify the doctors' omission to treat in the latter is very strong. We conclude that Hills' account, as it stands, does not accommodate even many or most pairs of cases involving harm-allowing (rather than harm-doing).

### **8. The "Moral Generalization" Solution**

Leaving past proposals aside, a defender of DDE might be drawn to a suggestion of Hart (1968).<sup>18</sup> The thought is that events or states of affairs are too close to each other to permit the possibility of intending the first without intending the second when they are "immediately and invariably" connected, in the way that someone's violently striking a glass is connected to its shattering.<sup>19</sup> Thus it might be argued that in Craniotomy, the crushing of a child's skull is immediately and invariably connected with its dying; in Sophisticated Terror Bomber, impacting enemy civilians with bombs is immediately and invariably connected with their dying (or suffering serious injury); in Guinea Pig, the disease running its course is immediately and invariably connected with the patients' suffering harm, and so on. The main problem with this suggestion is that hysterectomies performed on pregnant women are also immediately and invariably connected with the



death of their fetuses; that strategic bombing of military targets in densely populated areas is also immediately and invariably connected with the death of innocent civilians; and the failure to treat stubborn disease cases in order to save the greater number is also immediately and invariably connected with the patients' suffering harm.<sup>20</sup> So the "immediate and invariable connection" criterion does not enable us to distinguish morally between Craniotomy and Hysterectomy, between Sophisticated Terror Bomber and Strategic Bomber, or between Guinea Pig and Direction of Resources.

A more sophisticated variant of Hart's central idea has been defended by Wedgwood (2011). According to Wedgwood, the "central claim" of DDE is "that there is normally a stronger reason against an act if the act has a bad state of affairs as one of its intended effects than if it has that bad state of affairs as one of its unintended effects" (2011, 386). Wedgwood's account of intention, like Hills', borrows significantly from Bratman (1987) (and also Mele (2000)—see 2011, fn. 5, 389). To execute an intention, Wedgwood tells us, is to behave "in a way that is *guided* and *controlled* by that intention" (2011, 388). This conception of intention, as we have seen, is quite thin. So, for example, in Large Man, the bystander's intention is that the large man collide with the runaway trolley, and everything the bystander does in the service of saving the five is guided and controlled by that intention. As Wedgwood puts it, the guiding thought is not, "Let's make sure that the large man gets killed", but rather "Let's make sure that the large man collides with the trolley". Thus, he says, "proponents of the DDE must accept that in [Large Man], [the bystander] intend[s] the collision but *not* the death" (2011, 394). Wedgwood's proposed solution to the closeness problem, then, unlike Fitzpatrick's, does not appeal to some principle according to which if the obtaining of S1 is sufficiently

close (in some to-be-defined sense) to the obtaining of S2, then intending that S1 obtain entails intending that S2 obtain.

Wedgwood's solution to the closeness problem derives from his account of what makes states of affairs bad. Some states of affairs are *intrinsically* bad, while others are *extrinsically* bad. A state of affairs is extrinsically bad when "[i]t is bad only because of certain additional attendant circumstances"; it is intrinsically bad when it is bad but not extrinsically so (2011, 395). Someone's being in pain, then, is intrinsically bad. By contrast, a man's death is extrinsically bad, for "it is bad only because of certain additional attendant circumstances (such as the fact that the man wishes to go on living, or that his death deprives him of a reasonably good life, or the like)" (2011, 395). But, as Wedgwood recognizes, because we cannot say that the bystander in Large Man intends the large man's *death*, it is not possible to argue that it is the extrinsic badness of this death that explains why the bystander must be interpreted as intending something bad, and hence as subject to DDE. So Wedgwood searches for extrinsically bad states of affairs other than the large man's death that might serve as the relevant agents' objects of intention in such cases.

A state of affairs S counts as extrinsically bad, according to Wedgwood, whenever it is "bad news" in the sense that "Oh no, how awful! That sounds terrible!" would be an appropriate reaction to hearing of S's obtaining. And whether this kind of reaction is appropriate is determined by whether a virtuous agent would so react on the basis of her knowledge of certain true *ceteris paribus* moral generalizations, i.e., generalizations of the form "other things being equal, and under normal conditions, when S obtains, something bad (such as serious injury or death) obtains". So, for example, a

man's being struck by a trolley is "bad news", and hence extrinsically bad, because it would be appropriate for a virtuous person to react negatively to hearing of its obtaining, on the basis of her knowledge of the following *ceteris paribus* moral generalization: "[O]ther things being equal, and under normal conditions, when a collision between a person and a runaway trolley occurs, the person suffers serious injury or even death as a result" (2011, 396). But if a man's being struck by a trolley is an extrinsically bad state of affairs, then it is a bad state of affairs, and hence DDE will speak against any act that has that state of affairs as one of its intended effects.

Wedgwood's way of reconciling DDE and (Large Man, Trolley), then, is to say that even though the Large Man bystander does not intend the bad state of affairs consisting of the large man's *death*, the bystander *does* intend the bad state of affairs consisting of the large man's being struck by the trolley. By contrast, the Trolley bystander intends no more than that the trolley be shunted onto the side track; and *that* state of affairs, considered by itself independently of its consequences, is neither intrinsically nor extrinsically bad, for there is no true *ceteris paribus* moral generalization of the form, "other things being equal, and under normal conditions, when a trolley is shunted onto a side track, someone suffers either serious injury or death as a result".

One of the virtues of Wedgwood's view is that it appears to translate easily to other pairs of cases, such as (Sophisticated Terror Bomber, Strategic Bomber) and (Craniotomy, Hysterectomy). In Sophisticated Terror Bomber, the bomber intends enemy civilians' being impacted by the bombs in such a way as to appear dead. But what the bomber intends here, according to Wedgwood, is an extrinsically bad state of affairs, because it would be appropriate for a virtuous person, hearing of its obtaining, to react

negatively on the basis of her knowledge that, other things being equal, and under normal conditions, when people are impacted by bombs in such a way as to appear dead, they suffer serious injury or death as a result. Similarly, in Craniotomy, the doctor intends the fetus's skull's being crushed. And what the Craniotomy doctor intends here is an extrinsically bad state of affairs, because it would be appropriate for a virtuous person, hearing of its obtaining, to react negatively on the basis of her knowledge that, other things being equal, and under normal conditions, when the skull of a fetus is crushed, it suffers serious injury or death as a result.

But there are serious, and ultimately intractable, problems with Wedgwood's proposal. In the first place, Wedgwood's solution fails to account for our intuitions in Strategic Bomber and Hysterectomy. In Strategic Bomber, the bomber intends the munitions factory's being destroyed. So the important question here for Wedgwood is whether there are any true, *ceteris paribus* moral generalizations entailing the extrinsic badness of what the bomber intends. Leaving aside the people who tend to work *in* munitions factories (including janitorial staff who clean, and security personnel who police, such factories after hours), the fact is that most munitions factories, in large part because it would be hugely inefficient to locate them some distance away from the people who work in them, are built in urban areas that are densely populated with civilians. Given these facts, then, it is simply *true* that other things being equal, and under normal conditions, when munitions factories are destroyed, civilians are seriously injured or killed. The reason is that bombs that destroy munitions factories produce fallout, further explosions, and fires that maim and kill people living in nearby houses, working in nearby shops, and walking on nearby streets. In Hysterectomy, the doctor intends the

fetus's being removed from its mother's body. So the important question for Wedgwood is whether there are any true, *ceteris paribus* moral generalizations entailing the extrinsic badness of what the doctor intends. And here the situation is even less complicated. For it is true (indeed, obvious) that other things being equal, and under normal conditions, when an unviable fetus is removed from its mother's body, it dies. It follows directly that what the bomber in Strategic Bomber intends and what the doctor in Hysterectomy intends are both (extrinsically) bad, and hence that Wedgwood's version of DDE treats agency in these cases as no less disfavored than it treats agency in Sophisticated Terror Bomber and Craniotomy. This is both an unwelcome result for Wedgwood, and also deeply counterintuitive.

Wedgwood's proposal also entails counterintuitive consequences in other cases. Imagine a variant of Large Man (call it "Padded Large Man") in which the large man just happens to be padded to the hilt, in such a way that if he were pushed into the path of the trolley, his body and padding would stop the trolley and he himself would remain completely unharmed. It is important here that it is not normal for large men to be padded in this way, and thus it is true that other things being equal, and under normal conditions, when large men are struck by runaway trolleys, they are seriously injured or killed. And it follows from this truth and from Wedgwood's account of extrinsic badness that it is extrinsically bad for the large man to be struck by the trolley in Padded Large Man. But this is counterintuitive. If the large man is indeed padded to the hilt and will be unharmed if the trolley hits him, then his being struck by the trolley is neither intrinsically nor extrinsically bad.<sup>21</sup> And, indeed, in Padded Large Man it seems morally completely unproblematic to push the large man into the path of the trolley if this is the

only available means of saving the five. But given his theory of extrinsic badness, Wedgwood's version of DDE speaks against doing so.

Padded Large Man is a case in which Wedgwood's theory classifies a state of affairs as extrinsically bad when it is in fact neither good nor bad (or just good). There are other cases in which Wedgwood's theory classifies a state of affairs as neither good nor bad (or good) when it is in fact extrinsically bad. Surprisingly, Wedgwood himself discusses a case of this sort (call it "Spy"):

[S]uppose that I intend to catch your attention, so that you will look up and notice that I am in the room, because this is the agreed signal that will alert a spy who is observing us, thereby enabling the spy to defuse a bomb that would otherwise kill many people. (Suppose that I know that there is no other way in which the bomb can be defused if I do not catch your attention in this way.) I have also just found out that you have a bizarre condition so that when I catch your attention in this way, it will result in your instant death. (2011, 398)

To our eyes, given your bizarre condition, the state of affairs of your attention's being caught in this way is extrinsically bad because it immediately leads to your death. But, as Wedgwood explains, his own theory of extrinsic badness entails the opposite conclusion; for "[i]n this case, my intention is not a bad intention: there is no true *ceteris paribus* moral generalization to the effect that under normal conditions, catching a person's attention results in any serious injury or harm" (2011, 398). Although it does not strike Wedgwood this way, this strikes us as a particularly large bullet to bite. Without

intending to do so, Wedgwood has articulated a rather clear counterexample to his own theory of extrinsic badness.

Beyond facing problems related to its entailing counterintuitive results, Wedgwood's proposal is subject to the criticism that its classification of states of affairs as good or bad is completely arbitrary. Consider again whether a padded large man's being struck by a runaway trolley is extrinsically bad. According to Wedgwood, this question is answered in the affirmative if and only if the relevant *ceteris paribus* moral generalization governing the situation is true. But it turns out that there are at least two potentially relevant *ceteris paribus* moral generalizations in this case, namely (MG1) and (MG2):

(MG1) Other things being equal, and under normal conditions, when *large men* are struck by runaway trolleys, they suffer serious injury or death.

(MG2) Other things being equal, and under normal conditions, when *padded large men* are struck by runaway trolleys, they suffer serious injury or death.

As it happens, MG1 is true, but MG2 is false. So if we take MG1 to be relevant to the question of extrinsic badness, then the padded large man's being struck by the runaway trolley is extrinsically bad. But if we take MG2 to be relevant to the question of extrinsic badness, then the padded large man's being struck by the runaway trolley is not extrinsically bad. The problem is that Wedgwood's theory gives us no way to decide which of these two moral generalizations determines the question of extrinsic badness.

The choice between them seems arbitrary, and hence Wedgwood's classification of states of affairs as good or bad (or neither) is arbitrary too.

Indeed, the arbitrariness problem mushrooms. For there are many different levels at which states of affairs can be described. Consider the original Large Man case, in which the large man is not padded. The large man's being struck by a trolley *can* be described as a large man's being struck by an *object*, as a large man's being *touched* in some way by a trolley, as a large man's being touched in some way by an object, as a large man's being related in some way to an object, and so on. But the large man's being struck by a trolley *can also* be described as a large man's being struck by a *three-ton* trolley, as a large man's being struck by a trolley *moving at 100 mph*, as a large man's being struck by a three-ton trolley moving at 100 mph, and so on. To determine whether a particular state of affairs is good or bad, then, Wedgwood must choose between far more than two potentially relevant *ceteris paribus* moral generalizations. There are in fact uncountably many potentially morally relevant generalizations for him to choose from. And depending on which he chooses, he will get very different answers. For example, on the one hand, it is true that other things being equal, and under normal conditions, when a large man is struck by a three-ton trolley moving at 110 mph, he is seriously injured or killed; but on the other hand, it is false that other things being equal, and under normal conditions, when a large man is touched in some way by an object, he is seriously injured or killed.

Wedgwood might reply that the best level of generality at which states of affairs should be described in order to determine whether they are extrinsically bad is fairly low. So, for example, he might point out that at all levels of description that are *less general*



than “the large man’s being struck by a trolley” (such as, “the large man’s being struck by a three-ton trolley” or “the large man’s being struck by a trolley moving at 100 mph”), the relevant *ceteris paribus* moral generalization turns out to be true. But leaving aside the fact that such a reply does not articulate a *reason* for fixing on relatively low levels of generality, there is also the problem that it does not mesh with Wedgwood’s choice of generality level in other cases. For example, in Spy, Wedgwood chooses to describe the relevant state of affairs as “catching a person’s attention”. But this description lies at a fairly high level of generality (at least, relative to the Spy scenario). To see why, imagine that the bizarre life-threatening condition in the case is called “Efil Syndrome”, and consider that the relevant state of affairs in Spy could also be described at the following lower level of generality: “catching the attention of a person with Efil Syndrome”. (And the level of description matters: whereas it is *false* that other things being equal, and under normal conditions, *catching a person’s attention* results in serious injury or death, it is *true* that other things being equal, and under normal conditions, *catching the attention of a person with Efil Syndrome* results in serious injury or death.) The upshot is that any reply offered on Wedgwood’s behalf that fixes the relevant level of descriptive generality as fairly low conflicts with the way in which Wedgwood analyzes cases such as Spy. Even more problematically, should Wedgwood give up his analysis of Spy by fixing on a lower level of descriptive generality, it remains true that his method of determining extrinsic goodness and badness is open to the objection that the level of descriptive generality can be fixed at whatever level turns out to be compatible with one’s own intuitions about cases. And this problem strikes us as intractable.

## 9. The Intend/Foresee Distinction and Two Rationales

Thus far, we have canvassed a significant number of interesting and ingenious attempts to solve the closeness problem, and have found them all wanting. For a whole host of different reasons, it seems exceedingly difficult to avoid the conclusion that the classical DDE does not discriminate against sophisticated terror bombers, craniotomy doctors, doctors who withhold treatment in order to learn more from the progression of a disease, or bystanders who push large men in front of runaway trolleys. Indeed, we think that our investigation provides strong evidence that such a conclusion is ultimately inescapable. The evidence, such as it is, therefore suggests that DDE, as classically understood, is false.

We believe that there is a deeper unifying reason for the common failure of these various attempts to preserve the classical formulation of DDE in terms of the distinction between intended and foreseen harm. To begin, it will be helpful to consider the two main rationales that have been offered in support of DDE. The first is that DDE rests on a duty not to treat others as means to our ends without their consent (call this the “using-as-a-means” rationale), and the second is that it rests on a duty to avoid aiming at evil (call this the “aiming-at-evil” rationale), defended by Nagel (1986, 181-82) among others.

Interestingly, those defending solutions to the closeness problem we have examined here who discuss a rationale of any kind point to the aiming-at-evil rationale and not to the using-as-a-means rationale. We believe that this is not a coincidence. For while an explanation that invokes the right not to be treated as a means without one’s consent fits nicely with paradigm cases like Terror Bomber and Strategic Bomber, it does not fit perfectly with the general distinction between intended and merely foreseen harm

in general. And examining particular pairs of cases can bring this out. For example, consider Hills' version of Craniotomy, in which it is part of the doctor's plan to ensure the death of the fetus out of a concern to avoid a slow painful death, or, we might suppose, to avoid a continued but limited and painful existence that would be worse than death. In this case, the death is in fact intended (by stipulation), but it is not itself a means. Going in the other direction, we can see that harm can be merely foreseen, and yet still a person is used without her consent, and in such a way that harm comes to her. For example, in Guinea Pig, the doctors who let the disease go untreated use the patient without her consent, even if, as we have argued, they do not intend the harm. In general, then, one can intend harm without one's using the individual without her consent to advance one's original plan, and one can fail to intend harm and yet still use the individual as a means without her consent. The distinction between intended and merely foreseen harm simply does not make for a good fit with the rationale that appeals to a duty to avoid using others as means without their consent.

This leaves the aiming-at-evil rationale. And superficially, at least, it makes for a better fit with the traditional distinction between intended and merely foreseen harm. Intending is just aiming of a certain sort, after all. But when we try to explicate the rationale in a way that makes it more precise, we find that it is lacking in plausibility. Let us look more closely at Nagel's defense of the aiming-at-evil rationale. His proposal is that intending harm is more difficult to justify than merely foreseeing harm because it consists in aiming at (in the sense of being guided by) evil, the essence of which is that it should repel us. Nagel's proposed rationale is agent-centered, while the using-as-a-means rationale is patient-centered: on Nagel's view, the wrong-making feature in

question lies in the nature of the agent's intention; on the using-as-a-means rationale, the wrong-making feature lies in the infringement of the patient's right not to be used as a means without her consent.

We believe that there are several reasons not to accept the rationale. First, there is something unintuitive about any agent-centered justification of DDE: agent-centered justifications seem inappropriately self-indulgent.<sup>22</sup> Second, defenders of the aiming-at-evil rationale are caught on the horns of a dilemma depending on how they choose to understand the nature of evil. On the one hand, evil might be treated not merely as something *bad*, but as something *wrong*. But if wrongness is part of the essence of evil, then it is circular to explain the wrongness of an action (or its tendency to be wrong) by advertent to the fact that, in performing the action, the relevant agent aims at something that is wrong. On the other hand, evil might be treated as something bad (or very bad). In that case, another problem looms. For it does not seem wrong in itself to aim at something very bad (such as great harm): for example, it does not seem wrong in itself to aim at harming people who were or are engaged in wrongful attacks on other people. And if one proposes to avoid this problem by stipulating that what is wrong in itself is not aiming at great harm, but rather aiming at great harm *to those who were and are not engaged in wrongful conduct*, then explanatory circularity beckons once more. Third, the aiming-at-evil rationale risks explanatory failure. The problem is that the rationale tells us, in essence, that the wrong-making feature of intending to harm is that it involves the intention to harm. But what needs explaining here is precisely why the intention to harm conduces to wrongness. It might be suggested here that the aiming-at-evil rationale is not the end of the story, and in particular that the rationale could be grounded in an even

deeper rationale. But in that case, the aiming-at-evil rationale would be incomplete and hence deficient by comparison with complete rationales such as the using-as-a-means account.<sup>23</sup>

We are then left with the following: a number of cases that are highly suggestive of some principle like DDE and the more general idea that intentions matter to moral permissibility; a plausible deep explanation in the form of the using-as-a-means rationale; the failure of a number of attempts to solve the closeness problem for DDE; and a mismatch between DDE and the most plausible rationale often invoked in support of it. In the face of these facts, one response would be to continue to seek a solution to the closeness problem while foregoing the most plausible rationale. But we believe that a better response is to embrace an alternative principle that allows us to preserve and explain most of our intuitions about cases, that is grounded in the plausible using-as-a-means rationale, and that accepts that intentions are relevant to permissibility, but not in the simple way that is suggested by DDE.

Our suggestion follows Quinn in distinguishing between harmful direct agency and harmful indirect agency. “Harmful direct agency is “agency in which harm comes to some victims, at least in part, from the agent’s deliberately involving them in something in order to further his purpose precisely by way of their being so involved”, while harmful indirect agency is “agency in which either nothing is in that way intended for the victims or what is so intended does not contribute to their harm” (1989, 343). With this distinction in hand, we suggest replacing DDE with this principle: In cases in which harm must come to some in order to achieve a good (and is the least costly of possible harms

necessary), the agent foresees the harm, and all other things are equal, a stronger case is needed to justify harmful direct agency than to justify equally harmful indirect agency.<sup>24</sup>

For our purposes here, it is important to see that this principle dovetails nicely with the using-as-a-means rationale at the same time as it accounts for the cases canvassed earlier. For example, in *Sophisticated Terror Bomber*, we find harmful direct agency, even if we agree that, strictly speaking, the bomber only intends for the victims to *appear* dead. For the bomber deliberately involves the victims in his plan to further his purposes in such a way that harm came to them. This use of his victims for his own ends explains why such harmful agency is importantly morally different from the harmful indirect agency that we find in *Strategic Bomber*. This explanation does not require making any of the very creative, but ultimately unworkable moves that we saw defenders of DDE make in order to show that the harm itself is intended. Yet we preserve a great deal of what is appealing about DDE, including its apparent cohesion with the motivating cases, its grounding in a plausible deep rationale, and its implications that intentions matter for permissibility. While this is not the place for a full defense of this alternative principle, we hope to have shown that defenders of DDE have good reason to abandon the attempt to solve the closeness problem, while still adhering to what we believe are its most important features.<sup>25</sup>

## BIBLIOGRAPHY

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

Bentham, Jeremy. (1780) *An Introduction to the Principles of Morals and Legislation*.

<http://www.econlib.org/library/Bentham/bnthPML1.html>

Bratman, Michael. (1987) *Intentions, Plans and Practical Reason*. Cambridge, MA:

Harvard University Press.

Cavanaugh, T. A. (2006) *Double-Effect Reasoning: Doing Good and Avoiding Evil*.

Oxford: Clarendon Press.

Chappell, Timothy. (2002) "Two Distinctions that do make a Difference: The

Action/Omission Distinction and the Principle of Double Effect," *Philosophy* 77:

211-233.

Delaney, Neil Francis. (2008) "Two Cheers for 'Closeness': Terror, Targeting and

Double Effect," *Philosophical Studies* 137: 335-367.

Finnis, John. (1991) "Intention and Side-Effects," in *Liability and Responsibility*, eds. R.

Frey and C. Morris (Cambridge: Cambridge University Press), 32-64.

Fischer, John Martin, Mark Ravizza, and David Copp. (1993) "Quinn on Double Effect: The Problem of 'Closeness'," *Ethics* 103: 707-725.

Fitzpatrick, William J. (2006) "The Intend/Foresee Distinction and the Problem of 'Closeness'," *Philosophical Studies* 128: 585-617.

Foot, Philippa. (1978) "The Problem of Abortion and the Doctrine of Double Effect," in *Virtues and Vices and Other Essays in Moral Philosophy* (Berkeley and Los Angeles: University of California Press), 19-32. Originally published in *Oxford Review* 5 (1967).

Foot, Philippa. (1985) "Morality, Action and Outcome," in *Morality and Objectivity*, ed. T. Honderich (London: Routledge & Kegan Paul), 23-38.

Hart, H. L. A. (1968) "Intention and Punishment," in *Punishment and Responsibility: Essays in the Philosophy of Law* (Oxford: Clarendon Press), 113-135.

Hills, Alison. (2007) "Intentions, Foreseen Consequences and the Doctrine of Double Effect," *Philosophical Studies* 133: 257-283.

Kamm, Frances Myrna. (1992) "Non-consequentialism, the Person as End-in-Itself, and the Significance of Status," *Philosophy and Public Affairs* 21: 354-389.



Marquis, Donald B. (1991) "Four Versions of Double Effect," *Journal of Medicine and Philosophy* 16: 515-544.

Masek, Lawrence. (2010) "Intentions, Motives and the Doctrine of Double Effect," *The Philosophical Quarterly* 60: 567-585.

McMahan, Jeff. (1994) "Revising the Doctrine of Double Effect," *Journal of Applied Philosophy* 11: 201-212.

Mele, Al. (2000) "Goal-Directed Action: Teleological Explanations, Causal Theories, and Deviance," *Philosophical Perspectives* 14: 279-300.

Nagel, Thomas. (1986) *The View From Nowhere*. New York: Oxford University Press.

Nelkin, Dana Kay, and Samuel C. Rickless. (forthcoming) "Three Cheers for Double Effect," *Philosophy and Phenomenological Research*, published online, DOI: 10.1111/phpr.12002.

Pruss, Alexander R. (2012) "The Accomplishment of Plans: A New Version of the Principle of Double Effect," *Philosophical Studies*, published online, DOI: 10.1007/s11098-012-9925-4.

Quinn, Warren S. (1989) "Actions, Intentions, and Consequences: The Doctrine of

- Double Effect,” *Philosophy and Public Affairs* 18: 334-351.
- Scanlon, T. M. (2008) *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA.: Belknap Press of Harvard University Press.
- Smith, Ian A. (2007) “A New Defense of Quinn’s Principle of Double Effect,” *Journal of Social Philosophy* 38: 349-364.
- Wedgwood, Ralph. (2011) “Defending Double Effect,” *Ratio* 24: 384-401.
- Woodward, P. A. (1997) “The Importance of the Proportionality Condition to the Doctrine of Double Effect: A Response to Fischer, Ravizza, and Copp,” *Journal of Social Philosophy* 28: 140-152.

## NOTES

---

<sup>1</sup> Many of the ideas developed below grew out of a graduate seminar on non-consequentialist ethics that we co-taught at UCSD in winter quarter 2010. We would like to thank the participants in that seminar for their incisive questions and contributions: Craig Agule, Amy Berg, Nanhee Byrnes, Alex Marcellesi, Veronica Pear, and our colleague, Dick Arneson. We have also benefited greatly from the constructive comments and helpful suggestions we have received from David Brink, William Fitzpatrick, Matthew Hanser, and Derk Pereboom. This paper was conceived and written as a companion to Nelkin and Rickless (forthcoming).

<sup>2</sup> Or, more generally, there is a morally significant difference between intending bad states of affairs (as means or as ends) and merely foreseeing bad states of affairs.

<sup>3</sup> On some views, the genesis of the absolutist version of DDE begins either with Aquinas's theory of morally permissible self-defense or with nineteenth-century interpretations of passages from Romans 3:8 and Romans 6 (see Bennett 1995, 200). Thereafter, the absolutist version undergoes both development and simplification by a variety of theologians and philosophers. For helpful details, see Cavanaugh (2006, chapter 1). We believe, following a suggestion of Fischer, Ravizza, and Copp (1993, 707), that the moral relevance of the intend/foresee distinction is argued for as early as the Talmud, and we would argue further that the distinction receives further development in Jewish law thereafter, particularly in the elucidation of Sabbath prohibitions (the doctrine of "davar she'eino mitkhavein" – something not intended) to be found in Maimonides' *Mishneh Torah* and Joseph Karo's *Shulchan Aruch*. Some absolutist versions of DDE offer not merely a necessary condition for moral permissibility (namely, that one not intend harm as a means or as an end), but also offer a set of jointly sufficient conditions for permissibility when an action A produces two effects, one good (G) and one bad (B): (i) that A be good, or at least indifferent, in itself; (ii) that B not be intended as an end or as a means to achieving G; and (iii) that G be in some way proportional to B. We think there are counterexamples to all absolutist versions of DDE, both those that articulate necessary conditions for moral permissibility and those that articulate sufficient conditions for moral permissibility. In the first place, moral permissibility does not require the absence of an intention to harm: arguably, it is morally permissible for us to defend ourselves against a vicious aggressor by harming him as a means to prevent him from harming us. And in the second place, as Foot (1978, 29) has argued, conditions (i)-(iii) are not sufficient for moral permissibility: it is arguably impermissible to manufacture a gas (a process that is indifferent in itself) that is intended to save five people if the manufacturing process itself produces fumes that kill one person whom we are unable to move and do not intend to harm. We therefore consider non-absolutist versions of DDE to be theoretically stronger than their absolutist counterparts. We note here, however, that it does not matter for our purposes in this paper which version of DDE (absolutist or non-absolutist) is stronger. The problems we raise for the classical DDE apply to both versions of the theory in equal measure.

---

<sup>4</sup> Some theorists (e.g., Bentham (1780, IX.24), Hart (1968) and Foot (1978)) prefer to think of the intention/foresight distinction as a distinction between different kinds of intention. What we call “intention” these theorists refer to as “direct intention”, and what we call “mere foresight” they call “indirect intention” or “oblique intention”. The thought that states of affairs that are merely foreseen but not directly intended should still be classified as intended in some sense is backed by the thought that it seems disingenuous to claim that one does not intend *S in any sense* when one does something to bring about *S* with the full knowledge that *S* will occur. Still, we think that talk of “indirect” or “oblique” intention is misleading, because it seems possible in principle to foresee an event that one causes or allows without intending its occurrence (see Bennett (1995, 203), and the best theories of intention (see, e.g., Bratman (1987)) reflect this intuition. For the purposes of this paper, however, it does not matter whether what we call “merely foreseen” states of affairs are intended in some sense. Everything that we say about the intention/foresight distinction carries over, *mutatis mutandis*, to the direct/indirect or direct/oblique intention distinction.

<sup>5</sup> See Cavanaugh (2006), Chappell (2002), Delaney (2008), Finnis (1991), Fitzpatrick (2006), Foot (1985), Hills (2006), Masek (2010), and Wedgwood (2011), among others.

<sup>6</sup> Versions of the closeness problem appear in the Talmudic tradition (see note 3) and in Pascal’s *Provincial Letters* (Letter Seven). Foot (1978) raises the problem in its contemporary form, and the theorist who has done the most to develop it since then is Bennett (1995).

<sup>7</sup> See, in particular, Quinn (1989), who defends an alternative non-absolutist, intention-based principle according to which it is, other things being equal, more difficult to justify harmful direct agency (“agency in which harm comes to some victims, at least in part, from the agent’s deliberately involving them in something in order to further his purpose precisely by way of their being so involved”) than it is to justify harmful indirect agency (“agency in which either nothing is in that way intended for the victims or what is so intended does not contribute to their harm”) (1989, 343). Quinn’s view has been criticized by, among others, Kamm (1992), Fischer, Ravizza and Copp (1993), McMahan (1994), and Scanlon (2008). For defenses of (or responses to objections to) Quinn’s view (or views very similar to Quinn’s), see Woodward (1997), Smith (2007), and Nelkin and Rickless (2012). For an alternative principle similar in many ways to DDE that replaces the intensional concept of intention with the extensional concept of accomplishment, see Pruss (2012).

<sup>8</sup> The presentation of these cases is taken directly from Quinn (1989, 336).

<sup>9</sup> Here it should be noted that assumptions about the moral status of the fetus play an important role in (Hysterectomy, Craniotomy). Those who think that the fetus does not have much (or any) moral status (i.e., that it is not a bearer of rights, or a bearer of very weak rights; that it is not a member or full member of the moral community), will likely

---

find themselves evaluating both medical procedures as equally morally justifiable. It is only among those who think that human fetuses are persons (or near-persons) that one finds the intuition that the medical procedure in Craniotomy is, other things being equal, more difficult to justify than the medical procedure in Hysterectomy.

<sup>10</sup> This description is adapted from Bennett (1995, 210-11). The monicker “Sophisticated” is borrowed from Hills (2006, 264), who describes the relevant bomber as “philosophically sophisticated”.

<sup>11</sup> A variant of the Causal Necessitation proposal (according to which it is *knowledge* of causal necessitation between states of affairs S1 and S2 that makes it impossible to intend S1 without intending S2) appears in the Talmudic doctrine of “psik reisha” (“it will inevitably happen”—see note 3 and Fischer, Ravizza, and Copp (1993, 707)). Karo, in the *Shulchan Aruch*, recognizes that “psik reisha” requires further development, at least in the context of Sabbath law (in the form of what has come to be known as the distinction between “nicha lei”—“one is happy with the outcome”—and “lo nicha lei”—one is not happy with the outcome).

<sup>12</sup> This is consistent with the claim that agents’ knowledge can affect the moral permissibility of their conduct, a claim with which we agree. Here we simply reject the idea that the knowledge of the “*plain man*” is a factor in determining the permissibility of an action.

<sup>13</sup> Like Cavanaugh (2006, 115), Delaney (2008, 337-38), and Fitzpatrick (2006, 589-90), Masek (2010, 572) claims that the bomber in Sophisticated Terror Bomber “plans to make the civilians *apparently* dead by making them *actually* dead”. But this is a mistake. The bomber certainly does not have the death of the civilians as an *end*, and he does not believe that the death of the civilians is a state of affairs in the causal sequence that will result in his end, which is the enemy’s surrender. The relevant causal sequence is not: drop the bombs, which causes civilian deaths, which terrorizes the enemy population, which causes enemy troops to surrender. (This is the relevant causal sequence in the classic Terror Bomber case.) Rather, thanks to Bennett’s clever ‘sophistication’ wrinkle, the relevant causal sequence is: drop the bombs, which causes the bombs to impact civilian bodies, which causes the civilians to appear dead, which terrorizes the enemy population, which causes enemy troops to surrender. The whole point of the ‘sophistication’ wrinkle is that the bomber in Sophisticated Terror Bomber does not need the enemy civilians who are impacted by the bombs he drops to *be* dead as a means to his end of winning the war; all the ‘sophisticated’ bomber needs as a means to his ultimate end is that the enemy civilians *appear* to be dead. Indeed, if the enemy civilians *were* in fact dead but, for some strange reason (say, because they were still standing, eyes open, and making apparently goal-directed movements) *appeared to be alive*, the ‘sophisticated’ bomber would take the relevant part of his plan to have failed, rather than succeeded.

<sup>14</sup> One would think that Lily’s wrongdoing is not limited to *stealing*. For Lily surely *assaults* Otto by cutting him open, and then *kills* him by removing his vital organs. So

---

one would expect Masek to agree that there are more moral principles applying to Living Patient Transplant than the principle that one should not steal. This should not affect Masek's analysis of the case, inasmuch as it is fully compatible with his basic claim that Lily's wrongdoing is to be explained by moral principles unconnected to DDE.

<sup>15</sup> It is not clear, however, whether Fitzpatrick can help himself to accounts of the constitution relation offered by metaphysicians who are interested in explicating the nature of the relation between statues and statue-shaped pieces of clay, or between a ship-shaped aggregation of planks and the ship that the planks compose. For these accounts concern a relation that obtains between objects or aggregates of objects, not a relation that obtains between states of affairs. There is no intuitively obvious way of translating an account of material constitution into an account of a relation that obtains between states-of-affairs.

Perhaps one can say this much. If object O1 constitutes object O2, then, for any property F, the state of affairs *O1's being F* constitutes *O2's being F*. So if one had an account of *object-constitution*, then one would thereby have an account of *state-of-affairs-constitution*. But this is not sufficient for Fitzpatrick's purposes, because he needs an account of what makes, say, *O's being blown to bits* constitute *O's being killed*. And a principle connecting *object-constitution* to *state-of-affairs-constitution* won't do the trick. What Fitzpatrick could use, here, is a principle connecting *property-constitution* to *state-of-affairs-constitution*: for his purposes, it would be useful to be able to say that given that the *property* of being blown to bits constitutes the *property* of being killed, it follows that *O's being blown to bits* constitutes *O's being killed*. Unfortunately, we do not know of any (defensible) principles of *property-constitution*.

<sup>16</sup> Fitzpatrick does not discuss Guinea Pig or Direction of Resources. But it is reasonable to assume that he would provide a similar analysis here too. In Direction of Resources, the doctors intend a large number of 'easy' patients' being treated for the disease, the (unintended) result of which is that a smaller number of 'stubborn' cases remain untreated; by contrast, in Guinea Pig, the doctors intend the 'stubborn' cases' remaining untreated and thereby (because of the natural constitution account) intend the 'stubborn' patients' suffering serious injury or death.

<sup>17</sup> And there are more questions. First, what is the relation of *partial* constitution, and how is the partial constitution relation related to the (whole) constitution relation? Second, is (knowledge of the existence of) a relation of partial (natural) constitution between S1 and S2 sufficient for its being impossible to intend that S1 obtain without intending that S2 obtain? If so, shouldn't the relevant natural constitution account be that A's intending that S1 obtain is sufficient for A's intending that S2 obtain whenever A knows that S1 *either partially or wholly* naturally constitutes S2? Third, could it really be true that an area's being destroyed *constitutes* its occupants' being destroyed? Isn't it true, instead, that an entire area's being destroyed *logically entails* its occupants' being destroyed, just as an entire area's being covered in paint *logically entails* its occupants' being covered in paint?

---

Pruss (2012, 18) offers two counterexamples to Fitzpatrick's claim that intending evil follows from "intending something that would be partly constitutive of an evil". The first is this: "Death is partly constituted by life: to die is to have lived and to live no more. Thus, the parents who intentionally conceive a child are intending something, life, that is partly constitutive of the evil of death." The second is this: "[S]uppose that you are a doctor who saves an innocent prisoner from a coma in a case where you foresee that the prisoner will be unjustly tortured but you reasonably judge that the torture is not going to be so severe as to have made it advisable to leave the prisoner in a coma. You bring it about that the prisoner is conscious, and being conscious is a constitutive part of the prisoner's being tortured."

We do not find these counterexamples convincing, however. As Fitzpatrick makes plain, the constitution (or partial constitution) relation obtains between states of affairs, not between such states as life or death. The claim that life partially constitutes death makes sense if read as the claim that it is not possible to die unless one has lived. But this is not the kind of claim that Fitzpatrick takes to be relevant to the closeness problem. In order to be relevant, the claim that life partially constitutes death would need to be read as the claim that (the state of affairs of) X's being alive partially constitutes (the state of affairs of) X's being dead. But the latter claim just seems straightforwardly false. As for the claim that X's being conscious partially constitutes X's being tortured (a claim that *does* posit a relation between states of affairs), we think that Fitzpatrick would simply deny it on the grounds that X's being tortured *logically entails* X's being conscious, and it is impossible for S1 to (partially) constitute S2 if S2 logically entails S1.

<sup>18</sup> Hart considers the consequences of this suggestion, but does not adopt it as a way of saving DDE. Indeed, Hart is skeptical of DDE because he thinks the suggestion doesn't succeed in saving the doctrine.

<sup>19</sup> See Quinn (1989, 338-39). The rest of this paragraph is indebted to Quinn's discussion of Hart's suggestion.

<sup>20</sup> It might be argued that hysterectomies *generally* (unlike craniotomies *generally*), that strategic bombing *generally* (unlike terror bombing *generally*), and that withholding treatment of stubborn cases *generally* (unlike withholding treatment of stubborn cases in order to watch a devastating disease progress *generally*) are not immediately and invariably connected with anyone's suffering harm. But it could happen that (for various complex sociological, psychological, or medical reasons) most hysterectomies were performed on pregnant women, and it may already be the case that most strategic bombing involves dropping bombs on densely populated areas, and so on. So, as Quinn (1989, 338) rightly points out, it is *ad hoc* to focus on the type "hysterectomy" instead of the type "hysterectomy performed on a pregnant woman", *ad hoc* to focus on the type "strategic bombing" instead of the type "strategic bombing of a factory in a densely populated area", and so on. Wedgwood (2011) provides something approaching an elaboration of Hart's suggestion. For discussion, see section 8 below.

---

<sup>21</sup> If one worries about the trauma that might be experienced by the large man if he were unaware of his padding or afraid of falling onto trolley tracks, we can alter the case to make him aware that his padding will leave him unharmed in case of a collision with a trolley or to make him sedated.

<sup>22</sup> We thank an anonymous reviewer of Nelkin and Rickless (forthcoming) for helping us to see this point.

<sup>23</sup> For further development of these points, see Nelkin and Rickless (forthcoming).

<sup>24</sup> This is inspired by (but a modified version of) Quinn's own suggested principle: All other things equal, a stronger case is needed to justify harmful direct agency than to justify equally harmful indirect agency (1989, 344).

<sup>25</sup> We offer a direct defense of this principle in Nelkin and Rickless (forthcoming).