

University of Pennsylvania Carey Law School

Penn Carey Law: Legal Scholarship Repository

Articles

Faculty Works

1-1-2015

Intentional Content and Non-Combatant Immunity: When Has One Intentionally Killed a Non-Combatant?

Kimberly Ferzan

University of Pennsylvania Carey Law School, kferzan@law.upenn.edu

Follow this and additional works at: https://scholarship.law.upenn.edu/faculty_articles

Recommended Citation

4 Methode: Analytic Persp. 88 (2015)

This Article is brought to you for free and open access by the Faculty Works at Penn Carey Law: Legal Scholarship Repository. It has been accepted for inclusion in Articles by an authorized administrator of Penn Carey Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.

Intentional Content and Non-Combatant Immunity: When Has One Intentionally Killed a Non-Combatant?

Kimberly Kessler Ferzan

Certain types of people are impermissible targets in war. For instance, under current norms, it is impermissible to target non-combatants. Just war theory revisionists, such as Jeff McMahan and Cecile Fabre, would alter who may be targeted during war, focusing more on the individual morality of self-defense to determine who the permissible targets are (McMahan 2009a, Fabre 2009). Under such views, some non-combatants are permissible targets and some combatants are not. Still, whether one adopts the traditional view or the revisionist one, the bottom line is that some people are impermissible targets.

At the moment, non-combatant immunity from attack is a “fundamental tenet of just war theory” and this idea “underpins many of the *just in bello* rules.” (Frowe 2011, p. 6). Just war theory continues to hold true to the Doctrine of Double Effect (DDE), thus maintaining that although it is impermissible to intentionally kill civilians, one may justify the knowing killing of civilians by the good consequences that follow.¹ Under this analysis, it is permissible to tactically bomb a munitions factory, knowing that it will kill nearby schoolchildren, but it is impermissible to intentionally kill the schoolchildren (to terrorize the population to win the war). Legally, intentionally killing non-combatants is a war crime (Frowe 2011, p. 104).

This non-combatant immunity seems to presuppose that we can tell the difference between combatants and non-combatants. At one time, an able bodied man would be a clear and permissible target. Now, however, women may be targets and many men may not. Hence, a soldier may be uncertain whether the individual he sees is liable to be killed in war. Or, he may have to engage in an action, such as dropping a bomb, where he knows that there may be some non-combatants amongst the combatants.

Recently, just war theorists have asked the question of whether one is intentionally killing a non-combatant if one is killing *this person* and there is a *risk* that *this person* is a *non-combatant* (McMahan 2011, Haque 2012). Whereas Adil Haque argues that belief to a high probability does lead to the conclusion that one is intentionally killing a non-combatant, Jeff McMahan maintains the opposite.

The question of whether uncertainty cases are cases of intention turns on intentional content. Neither Haque nor McMahan supply such arguments. In this paper, I will first set forth my previous arguments about the representational content of intentions. Then, I will address McMahan and Haque’s arguments, while drawing relevant distinctions along the way. Having found their arguments lacking, I will turn directly to the question of when belief that there is a probability that a circumstance exists

¹ I will use civilian and non-combatant interchangeably for our purposes, though some civilians participate in wars in ways that make them combatants.

figures into the representational content of an intention. I will also consider Gideon Yaffe's view that intentions can include circumstances to which one is reckless. I will argue that intentions, understood as the mental state that causally guides and explains actions, do not include known uncertainties.

Of course, this analysis of representational content is only the beginning of the question for just war theorists. To my mind, the DDE is both too weak and too strong. It is too weak because we need to ask a different question when we are killing under uncertainty. What we need to do is to settle on the degree of certainty required for determining that there is liability to defensive force. The question is whether the person is innocent or whether he has forfeited rights; that is, whether he is liable to be killed. Just as we have proof beyond a reasonable doubt for punishment, we must settle on the point at which one may permissibly harm another when one is uncertain whether he is an innocent person or a culpable aggressor. If the person finds his belief about the target meets the threshold for proof, then one may treat the target as if he is liable to force. The DDE itself does not supply the answer to this question. We need a further argument for what the standard of proof should be.

In addition, elsewhere, I have maintained that the DDE is too strong (Alexander and Ferzan 2009, p. 99). I think that it is permissible to kill innocent people so long as we don't use them. That is, the constraint on opportunistic killing is the correct question, not the question of whether the killing is intentional. Moreover, because, as I will argue below, intentionally killing someone one knows is a non-combatant is to intentionally kill a non-combatant, we ought to question whether just war theory ought to condemn all such acts. (For an overview of the normative debates over the applicability of the DDE to war, see Frowe 2011, pp. 140-151, McMahan 2009b.)

Although I will gesture at these normative concerns about the DDE, again, my goal is not to press them here. Rather, my goal here is to settle the content of intentions. We cannot let our normative conclusions drive our analysis of representational content. To do so is to get the inquiry exactly backward.

I. The Scope of Our Intentions: A First Pass

What we intend is first and foremost what is motivationally significant to us. Intentions explain "why?" we perform an action. Following Michael Moore, we can say that intentions are mental states that mediate actions by (nondeviantly) causing actions, and most importantly, causing actions in a way that rationalizes them. (Moore 2013, p. 151-52).

Intentions are thus contrasted with effects that we know will follow from our intended actions, but that we do not desire. The question about what distinguishes an intention from a side-effect is well known. Placing drapes in the sunlight causes them to fade; speaking with a stutter causes annoyance; aiming at the munitions factory will blast the school; and so forth (Finnis 1991, p. 46). In each of these cases, we say that something may be a known side-effect without being intended. Indeed, the DDE takes for granted that we can distinguish known side-effects from intended consequences. If every known consequence was also intended, the distinction upon which the DDE relies would not exist.

Fortunately for the DDE, there is a distinction between these two. Known consequences do not cause or explain actions. We can also see the difference in the norms of rationality to which beliefs and intentions are subject. Intentions constrain reasoning in different ways than beliefs do. If one intends an action, one will screen out alternatives, engage in means-end reasoning, and be resistant to reconsideration (Bratman 1999, p. 141). These norms do not apply to beliefs. Conversely, stuck in equipoise, one may flip a coin and decide which of two paths to intend, but one cannot resolve beliefs in such a way (Harman 1991, pp. 63, 65). One could not, unsure whether her spouse is cheating on her, simply flip a coin to determine what to believe.

This distinction, however, is not the end of the matter. There are the tricky “closeness” cases. Does a scientist who intends to decapitate intend to kill? Can the terror bomber say actual death is just a known side-effect of making the victims “look dead?” (As imagined by Bennett (1980)). Elsewhere I have argued that contrary to conventional wisdom, the scope of an intention is not only those factors that are motivationally significant, but also those factors that are understood by the agent to be conceptually and empirically entailed by the factors that are motivationally significant (Ferzan 2008). That is, the actor’s understanding of *what she is doing* requires an understanding more robust than simply one thin linguistic description. To illustrate, someone who knows what it is to decapitate knows that what one is doing is killing. The one conceptually and empirically entails the other. The terror bomber does not know of a way to just make the victims “look dead.” He knows that bombs kill. The world we live in is one in which bombs kill and people die when their heads are removed from their bodies. Someone who did not know these things would have diminished rationality.²

Let me briefly explain what the account on offer is. It is an account of intentional content. If A intends to decapitate B, then “decapitating B” is A’s intentional content. But that content is not just words. Rather, that content must mean something to A. The content of the intention, then, is not only the description under which the intentional object is motivationally significant, but also all descriptions that inform the agent’s understanding of her act. A rational person would understand that in this world there is simply no way for a decapitated person to live. Hence, understanding what it means to decapitate B is understanding that one is killing B.

This connection is defeasible. It requires that the person be rational, understand basic physics and biology, and the like. Because we live in a world where people die when their heads are removed from their bodies, most people’s understanding of decapitation is that it is a form of killing. If someone did not know this, then (1) he would not believe it to be true and (2) he could certainly not intend it. But what matters is the agent’s understanding of the act. So, if conversely, the person falsely believes that

² Although Andrew Simester and I disagree about when circumstances fall within an intention, we agree about when results do. Simester sets forth his test for instantiations of meaning and expectation with which I concur: Some outcome Y is inseparable from X if, although there may exist worlds in which bringing about X without Y is possible, the worlds as the agent understands it admits of no such possibility. In cases of inseparability, the agent’s practical conception of the inevitable outcome of her behavior in bringing about X shall include Y—she cannot conceive of bringing about X alone. (Simester 1996, p. 459)

paper cuts immediately and necessarily cause death, then that person intends to kill when he intends to give a paper cut.

These cases are then distinguished from cases where one can understand one's action as excluding the side-effect. Bombing munitions factories can occur without killing schoolchildren. One could stutter without annoying others. Even if in an individual case a consequence is known, that consequence is not part of one's intentional content unless one cannot understand what one is doing without understanding that one is causing the known consequence.

There is a further question about closeness cases. How should we treat circumstances as opposed to results? Circumstances can have motivational significance, though we might think those cases are less frequent. If Alex intends to have sex with *this woman*, he may know she doesn't consent, but it takes a significantly more evil actor to be motivated by that fact. I may take a laptop because I want it, knowing it is yours, but it is less frequent for me to take the laptop *because* it is yours. Still, even if the cases wherein circumstances have motivational significance are rare, we must now ask the further question, which is whether one only acts intentionally vis-à-vis a circumstance when that circumstance is motivationally significant.

Again, it is simply not the case that the scope of our intentions are determined by the description that is motivationally significant. Other descriptions are equally part of the representational content. Consider a case where one decides to intentionally kill "Albert." One does not think one is going to kill a string of letters, specifically A-L-B-E-R-T. Rather A-L-B-E-R-T is a human being, a man, the guy over there, the person who stole one's girlfriend, and so on.

Consider the following explanation from Michael Luntley:

The fundamental insight [that drives the whole conception of sense and reference] is that if you could factor out grasp of the sense of a singular term from grasp of the sense of whole sentences, you would have no account of the rational power of the sense of the singular term. Thinking of an object is normative. To think of an object is to have your cognitive attitude to it subject to normative rational evaluation. The normativity of thought consists in the way a thought is systematically connected to others.

....

The suppositions that I have considered are all ways of revealing the way in which thought about an object must be sensitive to a cluster of thoughts that, as it were, provide the triangulation that fixes thought on a particular. You cannot, for example, demonstratively think about an object without having some idea of how it stands above, behind and to the side of other things, for if you did not have some idea about that, you would have no idea of its space-occupancy at all. (Luntley 1999, p. 236)

In other words, to have an understanding of killing A-L-B-E-R-T is to give A-L-B-E-R-T content. One thus intends what A-L-B-E-R-T means to one, *a meaning that will extend beyond motivational significance*. If one intends to kill A-L-B-E-R-T under the description of "that louse who cheated on me,"

one also intends to kill a human being, a man, a man wearing a red sweater, a man over there across the restaurant, and the other descriptions that give meaning to the object of one's intention. That is, the ascription of meaning given to an intentional object is not a one-dimensional word but rather the entire array of senses that one ascribes to that word.

Now, one might ask why it is that circumstance elements take on more known circumstances than result elements do. Why is it not just that one *knows* that one is killing a human being, while one *intends* to kill a louse? The reason is that when one acts prospectively one is aware that the other descriptions may or may not follow—it is not, except in closeness cases which are intended, part of the agent's understanding that these results are nomologically or conceptually required to follow. The school children could move and one would still understand that one is bombing the munition's factory. With a person or object on the other hand, one is acting on *this person/thing*. To understand that I am writing with *this pen* is to be writing with *this blue pen* is to be writing with *the blue pen my mom gave me for my birthday*. To think about the object I am using is to invoke what it is to me, which includes all the descriptions I (consciously or preconsciously) attribute to it.

Of course, this understanding of intentions, casting intentional content as broader than the description that is motivationally significant for the actor, fails to discriminate between actors whom we may think differ in culpability. Intention cannot pick out one characteristic as motivationally significant. Take two cases. In the first, A intentionally kills B because B is African-American. One might argue that intention appropriately captures this because B's race is motivationally significant. That would mean, then, that if A only knows but is not motivated by B's race, then A does not intentionally kill an African-American. But this does not work. For if intention only includes those facts that are motivationally significant then we would have a hard time convicting me for intentionally killing a human being if Albert's status qua human being is not motivationally significant for why I want to kill him. So, if motivational significance is what is morally relevant, intention is not the thing to mark it. (Ferzan 2008, p. 1151).

For purposes of just war theory, the recognition that intentions are broader than motivational significance may cast doubt on whether the DDE draws the correct line between permissible and impermissible conduct. My goal, however, is simply to get our understanding of intentions right. Intentions are mental states that mediate between one's beliefs and desires and one's action. (Moore 2013, p. 152). Although there is a description under which the intentional object is motivationally significant – and it is this description that causes and rationalizes the agent's action—the intentional object is not a thin linguistic description but content imbued with the meaning attributed to it by the actor.

II. Intentions, Uncertainty, and War: McMahan and Haque's Approaches

To this point, I have discussed when a known description is part of the actor's intention. Just war theorists are asking whether one intends a description to which one attributes only a probability. Let's begin with the solutions proposed by Jeff McMahan and Adil Haque. Both scholars are too quick

with their conclusions, but analyzing what they think, what they are missing, and how they go astray will be instructive.

A. McMahan

In *Who is Morally Liable to be Killed in War*, McMahan imagines the following three cases.

- (1) Suppose a pilot in a just war sees roughly 100 people concentrated in an open area. He sees that one is clearly a civilian but that the others are uniformed unjust combatants. The latter are on a mission vital to the success of their unjust cause and will not surrender. The pilot's only weapon is a bomb. If he drops it, it will kill all 100. Most people agree that if he drops it, he need not intend to kill the civilian. This is a typical example of proportionate_w 'collateral damage'. This could in principle be true even if the numbers were reversed. If a single unjust combatant would otherwise succeed in achieving his side's unjust cause, it would be proportionate_w for the pilot to bomb him even if he were surrounded by 99 people identifiable as civilians. The difference in numbers need not affect the pilot's intention. The killing of the 99 innocent civilians would be a proportionate_w side effect of avoiding defeat in a just war.
- (2) Next suppose that, in the original case in which the pilot sees 99 unjust combatants and one civilian, he circles back to drop his bomb, only to discover that the combatants have put a uniform on the civilian, so that he cannot be distinguished from them. This need not affect the pilot's intention. He can still drop his bomb without intending to kill the civilian, even though he cannot identify him. And, as before, this could be true even if the numbers were reversed.
- (3) Finally, suppose again that it is known that of the 100 people below, one is a civilian who has been forced to wear a uniform and thus cannot be identified. But the just combatants are a helicopter crew who have no bomb but only a gun that fires bullets. They kill all 100, one by one. They know that one of their shots will kill an innocent person. But they do not necessarily violate the constraint against intentionally killing an innocent person. Again the same could be true even if the numbers were reversed. (McMahan 2011, pp. 557-58).

McMahan resolves these cases as follows. The first case is a case where the pilot can distinguish the civilian but not discriminate in his attack. In (2), the pilot can neither distinguish nor discriminate. McMahan says "if, as most believe the pilot need not intend to kill the civilian in 1, he need not do so in 2 either" (McMahan 2011, p. 558). McMahan then claims that the difference between (2) and (3) is only epistemic, and that "[t]his difference does not necessitate a difference in intention" (McMahan 2011, p. 558).

I would parse these cases somewhat differently. In (1) and (2), the question is about when a result is intended. Hence, McMahan's (1) and (2) cases are cases in which the bomber knows he will hit a civilian, but also can realistically conceive of dropping the bomb without killing the civilian. Though it is true that *in this case* bombing these people will kill civilians, it is not the case that all targeted bombs necessarily kill civilians. McMahan is thus correct to conclude that the bomber who knows he will kill a civilian does not thereby intend to kill one.

McMahan is too quick to conclude that the move from (2) to (3) is just an epistemic difference. You are a shooting person who may be innocent. This is not simply squeamishness about our willingness to pull a trigger one hundred times as compared to dropping a single bomb. Rather by shifting the action, we shift from causing a result to engaging in an action where a particular circumstance exists or is *desired*. The reason these cases are tricky is because we don't *cause* circumstances. The way that intentions work with respect to results and circumstances is simply different, and thus, it is not simply a difference of acting under uncertainty once or one hundred times, as what one thinks one is doing when one is acting changes when the question is whether one is causing a result as compared to when the question is whether one is acting when a circumstance exists.

As I argued above, given how known circumstances inform the content of our intentions, if one intends to kill A, and one knows that A is a civilian, *then one intends to kill a civilian*. You can drop a bomb, knowing there is a civilian below without intending to kill a civilian, but you cannot shoot a man, whom you know to be a civilian, without intending to kill a civilian because the civilian status is part of what informs your understanding of the person upon whom you are acting.

McMahan's project is not to ask whether one intentionally kills a civilian when one knows the person he is killing is a civilian.³ He asks whether one intentionally kills a civilian when one knows there is a 1% chance he is a civilian and when one knows there is a 99% chance he is a civilian. Although his maneuver is not sufficient for us to conclude that one does not intend to kill a civilian in case (3), we are not compelled to conclude the opposite. The critical question remains: How should we think about cases where one is uncertain as to whether a particular description applies? Let us turn to Adil Haque's analysis of this question.

B. Haque

Adil Haque directly engages the question of killing under uncertainty (Haque 2012). To illustrate, Haque discusses a scenario used to train U.S. Marines for Iraq:

You are in a five-vehicle convoy moving...at 60 mph. As you pass under an overpass you observe an adult male, with a grenade-sized object in his hand, looking over the pedestrian railing above your lane. You cannot tell what's in the man's hand. What do you do? (Haque 2012, p. 84).

Haque says the Marine Corps "correctly instructs its forces" to "[s]tay read, observe the man on the bridge and prepare to fire if you observe hostile intent or hostile act" (Haque 2012, p. 84).

Many commentators and the International Committee for the Red Cross advocate a balancing approach when one does not know if one is under attack (Haque 2012, pp. 71-74). But Haque says the Balancing Approach would render it permissible to kill if you think he is 20% likely to be an insurgent if you will save four people. If nine Marines saved, then 10% certainty is enough. Haque calls these results "morally indefensible" (Haque 2012, pp. 71-74). Haque rejects such an approach, arguing that not only has no one articulated how to arrive at a level of certainty that takes all the military and

³ He does take this to be true, but thinks the issue is more complicated (McMahan 2011, p. 557).

humanitarian conditions into account, but also asserting that such any determination must account for “the moral asymmetries between killing and letting die and between intentionally and unintentionally killing civilians” (Haque 2012, p. 67).

To begin with, if you are almost certain that a person is a civilian, then—unless you are highly irrational—you will also believe that person is a civilian. Moreover, if you intentionally kill an individual believing the individual to be a civilian, and that individual is in fact a civilian, then you intentionally kill a civilian.... Since intentionally killing a civilian is much worse than unintentionally killing a civilian and unintentionally killing a civilian is substantially worse than allowing a fellow soldier to be killed, it follows that intentionally killing a civilian is far worse than allowing a fellow soldier to be killed. (Haque 2012, p. 89).

I won't focus on Haque's claims as to the killing/letting die distinction or to its applicability when soldiers have associative duties to each other. Rather, my goal will be to focus on the claim that the uncertain soldier can be held to have intentionally killed a civilian. Although Haque ultimately adopts an even more stringent test, he begins by sweeping instances where an agent believes a fact to a high probability into cases of intention. To do so, Haque suggests a fancy two-step. Let's take a look.

The first step is to rely on the claim, supported by the Model Penal Code, that as to an attendant circumstance, purpose is satisfied when the condition is known. The actor need not act because of the attendant circumstance (Haque 2012, p. 91). Here, Haque finds substantial support among the law and legal theorists.

The second step is more controversial. Haque argues first that we cannot rely on recklessness as the appropriate question for drawing the boundary between permissible and impermissible action in instances of uncertainty. Haque maintains:

[A] recklessness inquiry would lead us back to the Balancing Approach we have already rejected. Even a recklessness inquiry that reflects the moral asymmetry between killing and letting die would still entail that it is permissible to intentionally kill an individual who is almost certainly a civilian if doing so might (but almost certainly will not) prevent a substantially greater (but not far greater) number of soldiers from being killed. (Haque 2012, p. 93).

Having rejected recklessness, Haque maintains that there are some circumstances in which we might deem someone who acted under uncertainty as acting knowingly, specifically willful blindness cases. Haque maintains that when a high probability in a fact exists, it “seems morally equivalent to an affirmative belief that the fact exists.” (Haque 2012, p. 94). Thus, because under the Model Penal Code, purpose as to an attendant circumstance is satisfied by knowledge, and because Haque maintains that belief to a high probability is “morally equivalent” to knowledge, then purpose as to an attendant circumstance includes belief to a high probability that a circumstance exists (Haque 2012, pp. 93-94). Because one cannot believe that a target is a combatant in the face of a belief that he probably is a non-combatant, Haque concludes that withholding judgment – that is, not forming the belief that the target is a non-combatant—should be treated as morally comparable to having the flat out belief. (Haque 2012, p. 94).

With the knowledge and willful blindness moves in place, Haque goes on to argue for an even stronger, objective threshold. Haque begins by correctly noting that the moral wrong of killing civilians is simply the moral wrong of killing human beings – the status of civilian does not make the killing more wrong (Haque 2012, p. 94). Or, to put the point another way, the question is when the human being forfeits his right, a forfeiture that can occur when he is a combatant. Haque then argues that only when the person truly is a combatant is one justified; if one reasonably believes the person is a combatant, then one is excused (Haque 2012, p. 97).

To my mind, Haque is asking the right question in the wrong way. Notice that the question we are after is a question we ask in other contexts. Under conditions of uncertainty, when we are adjudicating someone's rights, what probabilities must we assign before we are permitted to act? One clear case is the standard of proof for criminal conviction. Likewise, the question of whether someone is a combatant is equally a question of whether the actor is sufficiently convinced that the targeted person has forfeited his rights. Ultimately, objectivists such as Haque and myself, who believe the rightness of action turns on facts, not reasonable beliefs, would want to say that even if the burden of proof is satisfied if the agent punishes or kills a person who is actually innocent or a civilian respectively, then the person is wronged. Although I wonder whether framing Haque's approach as one of reasonable belief as opposed to say, lacking a reasonable doubt, is perspicuous, he is certainly right to ask the questions about how the consequences of such actions (potentially saving many people) ought or ought not to figure into the determination of where we set the probability.

Where are we? We are operating under two separate moral constraints. The first moral constraint is that one may not intentionally kill innocent people, though one may kill them as a side effect. The second constraint is with respect to the degree of certainty that one must have that a target is liable before one may kill him. To further our inquiry into intentions, then, let me be clear that for our purposes, my goal is not to analyze whether I think the first moral constraint is plausible (I do not.) Nor is it to argue about the degree of certainty for the second constraint. (Though I take that to be an important and necessary question.) Rather, taking both of these moral constraints to be true, I am particularly interested in their border—I want to know where the line is such that the prohibition on intentionally killing a human being is triggered. Thus, where I want us to return is to Haque's discussion of how the willfully blind soldier is actually intentionally killing a civilian. If one suspects that there is a high probability of a fact, does one intend that fact and under what conditions?

III. Intentions and Uncertainty

A. Known and Uncertain Content

As I have argued to this point, the content of one's intention extends beyond the description that is motivationally significant to the actor. If I intend to kill Doug, I intend to kill him based on all the descriptions I use for him. That is, I may intend to kill him based on the description, "The man who made fun of me at the party" but I also intend to kill "the man," "the philosopher," "the person with the fantastic book on overcriminalization," "my commentator on this piece" and the like. That is, to

understand what it means to be killing not just a four-letter word “D-O-U-G” that intention must have content, and that content is determined by what D-O-U-G means to me.

What this analysis yields is that the Model Penal Code’s position with respect to purpose as to attendant circumstances is exactly right. That is, purpose is satisfied by a belief that that condition exists. One may intend to kill a philosopher, when one intends to kill that man and one knows that that man is a philosopher.

What happens, then, if I am uncertain that a description applies? Certainly, if that description, likely or unlikely, is what motivates me, then I act intentionally as to that description. If I intend to kill “the person who *may* write the best book on mistakes of law” because I want to stop him before he does write such a book, then my uncertainty as to whether it will be the best book does not undercut my intention. But what if this is just a possible description that is not motivational for me?

Here, it seems hard to see why such a description figures into the content of my intention. To this point, I have argued that other known descriptions are what gives the intentional content meaning. But the same is not true when the actor is uncertain as to whether to ascribe a particular description to his intentional object. If Joan goes to a restaurant, orders the lobster, and thinks there is a 20% chance that it is the most expensive thing on the menu, it is hard to see how her understanding of “ordering the lobster” is equally described as “ordering the most expensive thing on the menu.” After all, by recognizing that that description may not be true, Joan is denying these descriptions have equivalent senses to her.

We see this more clearly when we look at results. Recall that we rejected that intending “to bomb the munitions factory” is intending “to kill the schoolchildren.” The reason we rejected eliding these two was because even if the actor thinks it highly likely in an individual case – indeed even if she knows in an individual case – it is not part of the meaning that the agent ascribes to her intentional object. That is, the recognition that an act can be understood without a consequence is sufficient to create a gap between what the agent understands her act to be and what she knows she is doing on a given occasion.

Returning to probable circumstances then, the belief that something may be true of an object or person likewise means the agent does not understand that circumstance as including that description. For so long as the actor reserves judgment and does not form the belief that a description is true of the attendant circumstance, she does not equate their meaning and cannot be held to intend the description that she does not believe (full stop) to be true of her action.

Notice that Haque rejects this argument. He claims that if one believes x to likely be true, one cannot believe $\sim x$. Hence, one should not get the benefit of reserving judgment by hedging on whether x is true. According to Haque, by being barred by his rationality from believing $\sim x$, the actor ought to be treated as believing x .

There are two reasons to reject this claim. The first is that it focuses on how beliefs work, not how intentions do. But one can intend something that one believes highly improbable (say I shoot you

from far away) and one can intend not to do something one believes is highly likely (I can intend to be good about what I eat at the Christmas party but believe it is highly likely that I will head right for the desserts and eggnog.) So, it is not the case that if I believe x to be likely that I cannot intend \sim x. The second concern is Haque's argument is a moral one, not a metaphysical one. He wants to argue that both people are morally equivalent. This may be true for our determination of where to draw the line for the degree of confidence one must have that an individual is liable before one may harm him, but it cannot be that we use moral equivalency to slip unintended circumstances within the scope of intentions. Moral arguments lack metaphysical power.

B. Considering Haque's Model Penal Code Two-Step Analysis

What of Haque's creative employment of the Model Penal Code? Consider Haque's approach. He (rightly in my view) endorses the Model Penal Code approach that purpose as to an attendant circumstance includes belief that the circumstance exists. But then, although adopting the Model Penal Code's language, Haque argues not for the conceptual equivalence of belief to a high probability and knowledge, but for the *moral equivalence* of the two. This is to ask an altogether different question.

That is, we may think that it is as wrong to knowingly kill as to intentionally kill, or even that it is sometimes worse to act recklessly than knowingly. All of these arguments are beside the point, however. The question is, assuming that the DDE is correct to distinguish intention from knowledge, how is it that we determine intentional content? Moral equivalence is the wrong kind of argument.

We need to ask another question, though. Can we get from belief of a high probability that a circumstance exists to purpose as to that circumstance? Textually, if one were relying on the Code itself, one could not.⁴ The Code only allows high probability to substitute when knowledge is an element. On its face, knowledge is not an element in purpose cases *nor* is knowledge even given as a way of acting purposefully as to an attendant circumstance. Rather, the statutory language on purpose focuses on belief and awareness. The Code thus would not support the Haqueian two-step in an ordinary criminal case. (Model Penal Code, § 2.02).

Did the Code intend to say anything about the scope of intention? The Model Penal Code's only justification for treating high probability cases as knowledge cases is that we typically infer knowledge from such a belief. (Model Penal Code and Commentaries, p. 248). Therefore, the Code essentially creates an irrebutable presumption out of an evidentiary inference. Even if this inference is sufficiently common that the Model Penal Code's willful blindness test is unproblematic as a test of knowledge, it still leaves unresolved whether it is appropriate to employ this inference in cases in which we are dealing with the scope of an intention. The Model Penal Code does not provide us with a reason to think this is true.

Ultimately, we may want to say that one needs to believe an individual is liable to a high probability before one can do anything that would otherwise be rights violating. However, there is a significant difference between condemning an act because the person, who believes it 75% likely that a

⁴ I thank Peter Low for discussion and for raising the clear textual issues with Haque's argument.

target is a civilian, is thereby acting wrongly *by intending to kill a civilian* and a determination that one acts wrongly because one lacks the degree of confidence necessary to treat someone as if they are not a civilian.

C. Yaffe on Intentions

Recently, Gideon Yaffe has offered an account of intentions that would sometimes extend intentions to probable circumstances (Yaffe 2012). Building on his work in criminal attempts, in a recent piece on accomplice liability, Yaffe again reiterates the thought that “our intentions constitute commitments to conditions that they do not also commit us to promoting” (Yaffe 2012, p. 18). I wholeheartedly agree. Indeed, my paper *Beyond Intention* was aimed at specifically articulating how intentional content might include conditions that we do not intend to promote. I intend to kill Doug because I intend to do so under the description “the man who made fun of me” but I am equally committed to causing the death of the person “who wrote *Overcriminalization*” by forming the intention. My work has aimed to give an account of how and when other descriptions get in “the intention box” so to speak.⁵

Let me briefly set forth Yaffe’s general view. To Yaffe, we can determine if a mental state is an intention if it is accessible by certain norms of rationality (Yaffe 2010, p. 68). Among these norms are the commitments to non-reconsideration and non-complaint (Yaffe 2010, p. 148). These norms apply even if one is not committed to causing or promoting a particular circumstance or occurrence (Yaffe 2012, p. 18). To Yaffe, content (that, using my locution, is not motivationally significant) can be part of the agent’s intention in two additional classes of cases. First, an alternative description bleeds into the intention because it is salient such that the agent represents the intention as an intention under that description (Yaffe 2010, pp. 150-52). Or, second, if as a matter of logical necessity the intention to do one act entails the intention to do the other. (Yaffe 2010, pp. 89, 111, 120-21).

Salience is what gets uncertainties within the intention box. Yaffe seems to think it entirely plausible that when one acts under uncertainty, that description can be part of one’s intention:

it is perfectly possible for a person who intends to aid and is *merely reckless* with respect to the fact that the intended aid *aids a crime* to have an intention that constitutes a commitment of non-reconsideration with respect to the commission of the crime. This is particularly likely in cases in which the fact that the aid in question might contribute to the commission is particularly salient, as it is, for instance, when the probability is high, the intended aid is substantial, and the crime in question is severe. [When these conditions are present] it is much more likely that the defendant represents the occurrence of the crime in the intention on which he acts in giving aid, even if he is merely reckless with respect to the possibility that he will be aiding a crime. (Yaffe 2012, p. 29).

⁵ This indeed was how Michael Moore and Gideon Yaffe spoke about it at the Analytical Legal Philosophy meeting at Yale.

Let me raise a couple issues with Yaffe's general conception of intentional content before I turn to his argument in high probability cases. First, there is just a general worry that Yaffe uses these principles of rationality to determine whether there is an intention. However, Michael Moore clearly has it right. The fact that intentions are subject to such rules of rationality are contingent, not constitutive, facts about them (Moore 2013, p. 151). The take-away for our purposes is that we should be skeptical of a view that explains that something is intended, not by how it explains or rationalizes action, or how our minds provide meaning, but rather by some test of whether holding some view would require people not to complain or reconsider.

Relatedly, then, is that these norms may not themselves do the work we need them to do. There is rational pressure not to reconsider something in light of what one believed when one deliberated, and there is rational pressure not to complain about known consequences. Accordingly, these norms cannot themselves determine whether a circumstance or consequence is part of an intention. Even granting Yaffe that how these norms place pressure will be different, the point is that one cannot simply by fulfilling the litmus test of finding rational pressure conclude that the circumstance or consequence is intended. One needs an independent argument about how the item gets in the intention box.

Third, Yaffe and I disagree about how an understanding of an action or circumstance can get into the intention box. Return to the decapitation point. Yaffe's claim is that an intention to decapitate is only an intention to kill if either the actor represents the intention as an intention to kill or if as a matter of logical necessity the intention to decapitate entails a killing in every possible world. (Yaffe 2010, pp. 89, 111, 120-21). Because a decapitation admits of the logical possibility of not being a killing (if not a nomological possibility), the intention to decapitate is not an intention to kill. (Yaffe 2010, pp. 120-21). So, how can we prosecute someone who intends to decapitate for intending to kill? Because the mental state is *just as bad*, says Yaffe. (Yaffe 2010, p. 123). I would think this is a reductio of Yaffe's entire conceptual apparatus. If, *ceteris paribus*, intending to decapitate isn't intending to kill (when one understands that the person can't live without his head), then there is a profoundly flawed theory of intention at hand.

Now, let's return to salience and uncertain circumstances. Yaffe's thought seems to be that if something is very likely, one is very aware of it, and it is a rather big deal, then it is more likely that one will represent that circumstance/consequence within his intention. Before turning to the plausibility of the claim, consider its potential breadth. The strategic bomber who aims to bomb the munitions factory, knowing that the blast will kill the schoolchildren, is no doubt keenly aware of the fact that she is killing kids. Imagine that the strategic bomber is profoundly thoughtful and thereby troubled by it. Asked what she is off to do today, the bomber replies, "I am off to drop a bomb that will turn cute kids into bloody burned corpses." Does she now intend to kill the kids? I suppose it depends upon where we begin. I would start by saying that the core understanding of an intention holds – it is the facts/circumstances/etc. that are motivationally significant to the actor. It is why the actor acts. The strategic bomber acts – not to kill schoolchildren, a result she deplores – but to bomb the munitions factory. That core goal fixes the intentional content. I have then argued that the only other descriptions that figure into intentional content are those that have the same meaning to the actor – the same sense

– as the intentional object. Doug is a human being to me. Decapitating is killing. But to the strategic bomber her understanding of “bombing the munitions factory” does not conceptually or empirically entail that every bombing of a munitions factory kills school children.

Yaffe is open to more content bleed than this. In trying to parse when something that is not motivationally significant is still part of the agent’s intention, Yaffe puts on the table the commitments of non-reconsideration and non-complaint. That is, if A represents her munitions bombing as “killing the schoolchildren” and B represents his as “bombing the munitions factory” (though B knows it will kill the schoolchildren), Yaffe claims the two have different rational commitments. A can’t complain because that is what she intended. Moreover, in deciding whether to reconsider the bombing *because it will kill the schoolchildren*, Yaffe maintains that there is a difference between the rational pressure not to reconsider ϕ ing when we intend to ϕ and the rational pressure not to reconsider our intention when we intend to ψ but we deliberated that intending to ψ will cause consequence, ϕ , but ϕ does not figure in our intentions. There are rational pressures against reconsideration in both instances, but only the first instance contains rational pressure because of the content of the intention itself.

The additional content bleed is grounded not in what the person wants or what the person understands she wants, but what sort of person she is. Someone who cares about, deplores, begrudges a known side-effect might represent her intention as “off to kill the schoolchildren” because she gloomily accepts that that will be the consequence of her action, not because she is motivated by it. Hence, the more pessimistic and the more caring a person is, the more likely Yaffe will condemn the actor as intending the side-effect.

This may at first appear to be a normative argument.⁶ But it isn’t. It is the claim that descriptions may be salient to us, including descriptions of our own intentions, without those descriptions themselves being part of the intention. Consider the following example that happened just minutes before I started typing this. I am writing this paper. My goal is to get representational content right with respect to this question within just war theory. As I write this I believe that I need to address Yaffe’s views. Yaffe’s work is driving me nuts. Nuts. It is very hard philosophically; it is very technical; and I can’t resist the thought that it is somehow wrongheaded. I just need to figure out how. Moments ago, I got off the phone with a colleague, wherein I ended by saying, “Okay, I am off to do some more Gideon bashing.” Yup. That’s what I said. But even as I write this, I am not committed to bashing Yaffe, because that off-handed remark as a way of referring to the fact that I am getting back to *this paper* with *this topic* only potentially entails criticizing Yaffe. It is just that this particular potential argument is salient to me. But perhaps I will decide I shouldn’t bash Yaffe. Would I have irrationally reconsidered

⁶ For a normative argument against allowing the luck of how an intention is literally formulated from dictating moral conclusions, see Moore 2013, p. 169, Duff 2012, p. 425. Alex Sarch nicely articulates the worry:

[I]t is often going to be a matter of luck whether a given conditions makes it into the explicit content of one’s intention or not. In many cases, whether it does might depend on factors outside one’s control that are not reflective of one’s culpability—*e.g.* whether one is pressed for time, whether one is distracted or cognitively limited. Therefore, it does not seem that the question of whether a given condition has made it into the content of one’s intention will always signal a greater commitment to, or a greater willingness to tolerate, the criminal aims of the principals—as opposed to stemming from arbitrary factors beyond one’s control. (Sarch 2015, p. 16).

the very thing I intended to do? I think not because “Gideon bashing” is just a proxy I am using for what I am really intending to do: “write a well argued and correct view of representational content.” Deciding not to write this paper would be to reconsider what I am intending to do, but I would not be reconsidering it if I simply decide not to “bash Yaffe” which is just the description I casually picked out on the phone. The thought here is that what I use to “pick out” my intentional object, might not truly be what I intend. What I literally say may not be what I mean, and we are after the underlying work the intention is doing and not for the phrase the agent uses to pick it out.

One final argument to consider is Yaffe’s claim that probability cases can be recast as disjunctive intentions, thus allowing him to place probabilities within the individual’s representational content. He presents a hypothetical where the actor can either get \$5 now or take a box, which either has \$100 or marijuana. If it is marijuana, the actor will be fined \$50. The actor determines that the expected value calculation yields he should take the box. Yaffe then says, that the actor intends [to take the \$100 or marijuana box]. (Yaffe 2010, p. 155). Yaffe argues that because of the disjunctive the actor is committed to not complaining if his intention matches the world – that is, if the box has marijuana. (Yaffe 2010, p. 157). Gosh, if I took the box and it had marijuana, my reaction would be a series of expletives uttered in complaint for surely though I took the risk, I did not intend to take a box with marijuana. I intended to take a *box*, which I thought *might* have marijuana. The fact that one is reckless as to a potential circumstance and chooses to act in the face of it does not entail that one’s intention can be represented as intending x with circumstance p or ~p. Rather, one intends to act on x, with a belief as to a probability of p. One wonders what is left of recklessness if anytime someone deliberates about a risk (and thus it is salient), one has formed a disjunctive intention to perform the action as if the risk were a certainty. If acting under uncertainty is almost always (in relevant cases) acting intentionally, Yaffe has somehow created a test for intention that cannot distinguish among very different attitudes.

Ultimately, I remain persuaded by the view that intentions are mental states that cause and rationalize actions. The motivational significance is what makes intentions, intentions. The reason why anything else counts is only because the other descriptions are just one and the same description of the same object. But salience, probability, and importance do not synonymy make. Rather, these facts make it quite likely that the actor should strongly consider whether to act in the face of uncertainty.

Conclusion: On Morals and Metaphysics

Our morals can’t determine our metaphysics. So, too, our metaphysics should not determine our morals. As I have repeatedly reiterated, my goal here has not been to make the claim that the boundary that divides permissible from impermissible conduct in war ought to be governed by intention. Rather, making a determination about permissibility must turn on why it is that intention is thought to be important and whether our understanding of what intentions are maps onto their supposed importance. Typically, intentions seem to matter because we think they track the reasons that motivate the actor. Given that intentions extend beyond motivational significance, we should question whether they are the appropriate marker.

In addition, sometimes it is wrong to take a known risk. And it may be wrong even though one is not acting intentionally as to that risk. We need not claim that riskings are intendings in order to condemn riskings.

In this paper, I have sought to argue that though McMahan is too quick to conclude what is not intended; Haque is too quick to conclude what is. My aim has been not only to intercede into the debate within just war theory but also to explore the less examined idea that probable circumstances can be part of our intentions. With respect to just war theory, these arguments only matter if we are right to rely on the moral significance of intention. I suspect not.

References

- Alexander, L. & Ferzan, K.K., with Morse, S.J. (2009), *Crime and Culpability: A Theory of Criminal Law*, Cambridge University Press, Cambridge and New York.
- American Law Institute, Model Penal Code and Commentaries (1980-1985).
- Bennett, J. (1980), 'Morality and Consequences', available at <http://tannerlectures.utah.edu/documents/a-to-z/b/bennett81.pdf>
- Bratman, M.E. (1999), *Intention, Plans, and Practical Reason*, CSLI Publications, Stanford.
- Duff, R.A. (2012), 'Guiding Commitments and Criminal Liability for Attempts', *Criminal Law and Philosophy* **6**, 411-427.
- Fabre, C. (2009), 'Guns, Food, and Liability to Attack in War', *Ethics* **120**, 36-63.
- Ferzan, K.K. (2008), 'Beyond Intention', *Cardozo Law Review* **29**, 1147-1191.
- Finnis, J. (1991), Intention and Side-Effects, in F.G. Frey and C.W. Morris, eds, in 'Liability and Responsibility: Essays in Law and Morals', Cambridge University Press, pp. 32-94.
- Frowe, H. (2001), *The Ethics of War and Peace: An Introduction*, Routledge, Abingdon, Oxon and New York.
- Harman, G. (1999), *Reasoning, Meaning, and Mind*, Oxford University Press, Oxford and New York.
- Haque, A.A. (2012), 'Killing in the Fog of War', *Southern California Law Review* **86**, 63-116.
- Luntley, M. (1999), *Contemporary Philosophy of Thought: Truth, World, Content*, Blackwell, Oxford and Malden, Mass.
- McMahan, J. (2011), 'Who is Morally Liable to be Killed in War', *Analysis* **71**, 544-559.
- McMahan, J. (2009a), *Killing in War*, Oxford University Press, Oxford and New York.
- McMahan, J. (2009b), 'Intention, Permissibility, Terrorism, and War', *Philosophical Perspectives* **23** 345-72.
- Moore, M.S. (2013), 'Yaffe's Attempts', *Legal Theory* **19**, 136-77.
- Sarch, A.F. (2015), 'Condoning the Crime: The Elusive Mens Rea for Complicity', forthcoming *Loyola University Chicago Law Journal*, available at SSRN: <http://ssrn.com/abstract=2537578> or <http://dx.doi.org/10.2139/ssrn.2537578>

Simester, A.P. (1996), 'Moral Certainty and the Boundaries of Intentions', *Oxford Journal of Legal Studies* **16**, 445-469.

Yaffe, G. (2014), 'Intending to Aid', *Law and Philosophy* 33, 1-40.

Yaffe, G. (2010), *Attempts*, Oxford University Press, Oxford and New York.