Legal Theory, **10** (2004), 273–310. Printed in the United States of America
Published by Cambridge University Press 0361-6843/04 $12.00 + 00

CONDITIONAL INTENT AND *MENS REA*

Gideon Yaffe*
University of Southern California

There are many categories of action to which specific acts belong only if performed with some particular intention. Our commonsense concepts of types of action are sensitive to intent—think of the difference between lying and telling an untruth, for instance—but the law is replete with clear and unambiguous examples. Assault with intent to kill and possession of an illegal drug with intent to distribute are both much more serious crimes than mere assault and mere possession. A person is guilty of a crime of attempt—attempted murder, for instance, or attempted rape—only if that person had the intention to perform a crime. Under the federal carjacking law, an act of hijacking an automobile counts as carjacking only if performed with the intention to kill or inflict serious bodily harm on the driver of the car.[1] In all of these cases, the question of whether or not a particular defendant had the precise intention necessary for the crime can make a huge difference, often a difference of years in prison, but sometimes literally a difference of life or death; sometimes whether the crime is one for which the death penalty can be given turns solely on the question of whether or not the actor had the relevant intention.

Now, we ordinarily recognize a distinction between a conditional intention to act, which we might express with the words "I intend to A if X," and an unconditional intention to perform the same action. Some people intend to vote if the polls say the election is very close; some just intend to vote, period. This paper is concerned to answer the following question: Under what conditions does a conditional intention to act satisfy the intent requirement

*Portions of this material were presented at California State University at Northridge, the University of Michigan and the University of San Francisco. Audiences on all three of these occasions were very helpful. I owe special thanks to Mitch Berman, Michael Bratman, Rob Kar and Al Mele for providing detailed comments on earlier drafts. Helpful comments were also provided by Larry Alexander, Steve Darwall, David Dolinko, Steve Finlay, Brandon Fitelson, Barbara Herman, Pamela Hieronymi, Alexander Hughes, Aaron James, John Ku, Herbert Morris, Jeff Paris, Doug Portmore, Peter Railton, Gideon Rosen, Andrew Sabl, Scott Shapiro, David Shoemaker, David Stump, Judith Thomson, Greg Trianosky, Manuel Vargas, Kendall Walton, Gary Watson, Donald Wilson, Jessica Wilson, Takashi Yagisawa, and Paul Zeitz. Thanks, finally, to Jules Coleman for his encouragement.

1. 18 U.S.C. §2119 makes it a crime to "take a motor vehicle from another by force and violence or by intimidation with the intent to cause death or serious bodily harm."

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

of a given category of action, the *mens rea* of the action?[2] In fact, the courts have found themselves having to answer this question, for defendants sometimes have admitted to possessing a particular conditional intention while denying that they possess the corresponding unconditional intention and have gone on to argue that their conditional intention is not sufficient for *mens rea*.[3] For instance, in a recent Supreme Court case, Holloway v. United States,[4] which served as the test case for the federal carjacking law, this form of argument was employed. Holloway and an accomplice threatened to kill drivers of parked cars if they refused to hand over their car keys. Holloway argued that he and his accomplice had no intention to harm those who complied but only those who did not; and as it happened, all of his victims complied. Thus, claimed Holloway, his crimes were to be distinguished from the crimes of those who, for instance, hijack a car and driver with the intent to kill or maim the driver later in a secluded location, after the driver complies with the demand to relinquish control of the vehicle. In a split decision, the Supreme Court rejected the argument, claiming instead that for the purposes of the federal carjacking law, conditional intent to kill or injure is sufficient for the intent to kill or injure specified by the statute.[5]

In State v. Irwin,[6] by contrast, the North Carolina Court of Appeals accepted an argument with exactly the same structure. Confined in jail for

2. More than just intention is almost always included in the mens rea conditions of an act. A delusional person, for instance, might intend to kill while failing to possess the mens rea of murder. Our purposes here, however, do not require distinguishing between the overall mens rea conditions of an act and its intent requirements.

3. This form of argument was honored in State v. Morgan, 38 Am. Dec. 714 (N.C. 1842); Hairston v. State, 54 Miss. 689 (1877); Stroud v. State, 95 So. 738 (Miss. 1923); Craddock v. State, 37 So. 2d 778 (Miss. 1948); United States v. Rushlow, 2 U.S.M.C.A. 641 (1953 CMA); United States v. McLean, 11 C.M.R. 755 (1953 CMR); State v. Kinnemore, 295 N.E.2d 680 (Ohio Ct. App. 1972); State v. Irwin, 285 S.E.2d 345 (N.C. Ct. App. 1982); and United States v. Randolph, 93 F.3d 656 (9th Cir. 1996). Arguments of this form were rejected in Thompson v. State, 36 S.W. 265 (Tex. Crim. App. 1896); People v. Connors, 97 N.E. 643 (Ill. 1912); Beall v. State, 101 A.2d 233 (Md. Ct. App. 1953); Shaffer v. United States, 308 F.2d 654 (5th Cir. 1962); Commonwealth v. Richards, 293 N.E.2d 854 (Mass. 1973); United States v. Arrellano, 812 F.2d 1209 (1987 U.S. App.); People v. Vandelinder, 481 N.W.2d 787 (Mich. Ct. App. 1992); United States v. Anderson, 108 F.3d 478 (3d Cir. 1997); United States v. Norwood, 948 F. Supp. 374 (1996 U.S. Dist.); United States v. Lake, 972 F. Supp. 328 (1997 U.S. Dist.); United States v. Romero, 122 F.3d 1334 (1997 U.S. App.); and Holloway v. United States, 119 S. Ct. 966 (1998).

4. 119 S. Ct. 966 (1998).

5. The Supreme Court was responding in part to a decision made by the Court of Appeals for the Ninth Circuit in United States v. Randolph, 93 F.3d 656 (9th Cir. 1996). In that case, Randolph was one of a group of men who threatened to kill a woman if she refused to hand over her money and her car. She complied, and later, after Randolph had left believing the woman to have been released, his accomplices assaulted her. The lower court's ruling, overturned by the Ninth Circuit's decision, confounded Randolph's conditional intention to injure the woman if she did not comply with his demands with his accomplices' unconditional intention, on which they acted, to injure the woman. The Ninth Circuit's Court of Appeals argued that the only relevant question was whether or not Randolph's conditional intention was sufficient for intent to cause the woman serious injury; he was not to be implicated in his accomplices' independent plans to harm the woman. They then honored the form of argument under consideration here and overturned the lower court's conviction.

6. 285 S.E.2d 345 (N.C. Ct. App. 1982).

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

another offense, Irwin grabbed a woman visiting another prisoner and held a knife to her throat, threatening to kill her if he was not released. A jury convicted him of assault with intent to kill and he appealed on the grounds that he did not intend to kill the woman but intended only to kill her if his demands were not met. The Court of Appeals reduced his charge to one of assault without the requisite intention.

So the courts have had to respond to the following argument, offered by various defendants:

1. An agent is guilty of a crime of type C only if he had the unconditional intention to A.
2. Although I had the conditional intention to A if X, I lacked the unconditional intention to A.

Therefore I am not guilty of crime C.

Call this the "Mere Conditional Intent argument." Some courts have chosen to construe the argument as posing the question of whether the law's usage of the term "intention" in the definition of acts of type C is meant to apply to both unconditional and conditional intention or only to unconditional.[7] Courts interested in limiting the precedential scope of their decisions are often drawn to this strategy, for it allows them to claim that their decision concerns only the interpretation of the particular statute at issue and not the general question of whether or under what conditions conditional intention is as good as unconditional for the purposes of *mens rea*. The interesting philosophical question raised by the argument, however, does not concern the correct interpretation of any particular statute but, instead, the general question sidestepped by courts that adopt this approach.

There are, however, arguments employed by the courts that do address the question that interests us. The first section of this paper examines arguments that are intended to show that a person who has a conditional intention *under certain circumstances* also possesses the unconditional intention that is a crucial element of the crime. To adopt this approach is to distinguish between those cases in which the Mere Conditional Intent argument should be rejected and those in which it should be honored by appeal to the presence or absence of the special circumstances that are claimed to lead someone with a conditional intention to possess the forbidden unconditional intention. Thus, together with a claim to the effect that a particular defendant's circumstances are of the requisite sort, arguments of this kind can be employed to show that premise 2 is false. A court that employs an

7. Cf. United States v. Anderson, 108 F.3d 478 (3d Cir. 1997). The court also takes this approach, among others, in Holloway, 119 S. Ct. 966 (1998). To take this approach is to focus on the question of whether or not premise 1 is true. As we shall see, there is another question that one can take to be posed by premise 1.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

argument of this sort, however, admits (perhaps merely for the sake of argument) that a conditional intention is not sufficient by itself for *mens rea*; but since in such a case the court claims that the defendant possessed the relevant unconditional intention, the defendant is still guilty. Section I of this paper argues that such arguments are doomed to failure because there are no circumstances of the needed sort under which a conditional intention will become or necessarily cause an agent to have an unconditional intention.

Alternatively, a court can find grounds for distinguishing between cases in which the Mere Conditional Intent argument should be honored and those in which it should be rejected by specifying conditions under which a conditional intention is just as bad as an unconditional intention. A court that adopts this approach in rejecting the Mere Conditional Intent argument claims that premise 1 is false; such a court claims, that is, that under certain circumstances a conditional intention is enough for satisfaction of the crime's *mens rea* requirement while recognizing that the defendant lacked an unconditional intention. The American Law Institute adopts this approach in the Model Penal Code's guidelines for treating the Mere Conditional Intent argument. The Model Penal Code has had a strong influence on the view of conditional intention espoused by philosophers of criminal law and has also influenced judges' decisions; it is specifically invoked to support the majority's ruling in *Holloway*, for instance, and it is accepted verbatim in *Handbook of Criminal Law*,[8] a standard reference text.[9]

In Section II, the Model Penal Code's guidelines are shown to be unsatisfactory. However, rejecting the Model Penal Code's guidelines does not require rejecting the general strategy that the framers of the Code adopt. The rest of the paper is spent developing an alternative account of the conditions under which a conditional intention is just as bad as an unconditional, for the purposes of assessing *mens rea*.

Toward that end, Section III compares the sort of commitment to conduct involved in a conditional intention with the kind involved in intentions that are not conditional. There are various points of similarity and various points of difference corresponding to similarities and differences in the rationality constraints governing the deliberations of agents with unconditional and conditional intentions. Section IV returns to the question of how the courts ought to treat the Mere Conditional Intent argument. Through application of the results about the differences and similarities between conditional and unconditional intention developed in Section III, an alternative to the Model Penal Code's approach is proposed.

One important cautionary point: Intentions that are in various ways subject to conditions are not thereby conditional intentions in the sense of

8. W.R. LAFAVE & A.W. SCOTT, JR., HANDBOOK OF CRIMINAL LAW 200, 613 (1972).

9. *Cf.* Holloway; United States v. Arrellano, 812 F.2d 1209 (1987 U.S. App.); United States v. Anderson, 108 F.3d 478 (3d Cir. 1997); United States v. Norwood, 948 F. Supp. 374 (1996 U.S. Dist.); and United States v. Romero, 122 F.3d 1334 (1997 U.S. App.).

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

interest here. For instance, for almost every intention there is some condition such that if it were to obtain, the agent would not do the intended action. We almost never intend to do something *no matter what*. But this does not mean that every intention is conditional in the sense of interest here. I might unconditionally intend to go jogging in the morning despite the fact that I will not if it is raining. In cases such as this, the condition does not enter into the content of the intention.[10] However, not every intention that has a condition in its content is a conditional intention.[11] Compare someone who intends to shoot if he hears a noise—a person who has a conditional intention of the sort that interests us—to someone who intends to make true the conditional "If I hear a noise, I shoot." Although both of these people have an intention that includes a condition, they are not necessarily committed to the same set of actions. The latter person, for instance, might take himself to have done as he intends when he puts in earplugs. A theory of conditional intention—an account, that is, of what, precisely, conditional intentions are—would have to specify the precise manner in which a condition must be included in an intention's content for that intention to be conditional in the relevant sense.[12] Here I am not undertaking to produce such a theory; our purposes will not require anything more than what has just been said: An intention is conditional only if a condition is included in its content in a particular intuitive (but here unspecified) way.

10. Courts have sometimes gotten confused about this point. A British court in Regina v. Easom, 2 QB 315 (1971), for instance, accepted the Mere Conditional Intent argument when the evidence showed only that the defendant had an intention that was subject to conditions, not that he had a genuinely conditional intention. Easom picked up a woman's handbag, looked inside, and found there to be nothing of value in it, at which point he put it back where it was. He was charged with attempted theft—an offense that requires the intention to steal—and argued that he did not intend to steal the handbag, he only intended to steal it if it contained something of value. The court honored the argument and acquitted Easom. However, to note the condition governing Easom's intention was simply to note that he would abandon his intention to steal the handbag if he discovered he had no reason to act on it. This does not mean that he lacked the intention to steal the handbag but, rather, implies that he had it, or had an intention functionally equivalent to it.

In his dissenting opinion in Holloway, Justice Scalia seems to suggest that the majority is committed to treating every intention that is subject to conditions as though it were sufficient for mens rea. In a similar vein, although with the opposite import, in *Holloway v. United States: Conditional v. Unconditional Intent to Kill*, 51 Mercer L. Rev. 959 (2000), Michael Douglas Owens praises the court for its decision on the grounds that in light of the decision, defendants cannot be acquitted on the grounds that their intentions are subject to conditions. The issue is discussed with much greater sophistication in J.P.W. Cartwright, *Conditional Intentions*, 60 Phil. Stud. 233–255, 236–240 (1990). *See also* John Finnis, *On Conditional Intentions and Preparatory Intentions, in* Moral Truth and Moral Tradition: Essays in Honor of Peter Geach and Elizabeth Anscombe 163–176, 164 (L. Gormally, ed., 1994).

11. Jack W. Meiland, The Nature of Intention 15–25 (1970), proposes a test for determining whether or not a particular condition enters into the content of an intention in the special way that is distinctive of conditional intention. For discussion, *see* Finnis, *supra* note 10, at 165–167.

12. For discussion, *see, e.g.*, Donald Davidson, *Intending, in* Essays on Actions and Events 93–96 (1980).

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

I.

In *Holloway*, the majority proposes conditions under which anyone who has a conditional intention to A if X also has an unconditional intention to A, and then goes on to claim that Holloway satisfies those conditions. This is intended to support the majority's view that Holloway possesses the intent required for carjacking under the federal law. Writing for the majority, Justice Stevens says:

> [T]he question is whether a person who points a gun at a driver, having decided to pull the trigger if the driver does not comply with a demand for the car keys, possesses the intent, at that moment, to seriously harm the driver. In our view, the answer to that question does not depend on whether the driver immediately hands over the keys or what the offender decides to do after he gains control over the car. At the relevant moment, the offender plainly does have the forbidden intent. (Majority Opinion in *Holloway*)

Although there is some ambiguity about what exactly is being said here, the majority would appear to endorse the following description of the case: Holloway's plan for stealing the car includes shooting the driver if she does not hand over her keys; he has the relevant conditional intention. But the plan also includes doing what is necessary to get the driver to hand over her keys; that is, he is also committed to seeing to it that the condition involved in his conditional intention does not obtain. His chosen means for inducing her to hand over her keys is to form the *unconditional* intention to kill her, an intention that he informs her of by pointing a gun in her face. Under the majority's view, then, Holloway formed the intention that the crime requires—an unconditional intention to kill or injure—and then changed his mind after the driver handed over her keys. Further, he planned all along to change his mind, should she comply. But the fact that he changed his mind does not change the fact that he had the unconditional intention called for by the statute.

The point that the court is making here is intended to be general. The only facts of this particular case that the court takes to be relevant are those that show that Holloway intended to steal the car and that he threatened to kill the driver if she did not hand over her keys. But if the point is general, then it should be a conceptual matter that anyone, or at least any rational person, who adopts a plan like Holloway's and executes it through the issuance of a threat has the requisite unconditional intention. This is false, however. The majority is assuming that Holloway depended on the driver acting irrationally in handing over her keys. If the driver really believed that Holloway unconditionally intended to kill her, she would have no reason to hand over her keys. How would that improve her condition? In addition to being in the sights of a person intent on killing her, by giving him her keys she would also give up her best means of escape. Of course, sometimes we

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

do plan on other people not acting rationally and we act in order to exploit this tendency. A person might cause another to flinch by scaring her. In such a case, the manipulator does not believe that flinching is somehow the prudent response to fear or that the person who flinches believes this. Similarly, it is possible that Holloway thought that if he frightened the driver enough, she would give in to his demands even while lacking the belief that by giving in she would thereby improve her condition. If this is what he thought, then perhaps he did indeed have the requisite unconditional intention.

However, coercion of the sort that Holloway inflicted on his victim is not necessarily like this. If Holloway assumed that the driver would make the rational, prudent decision, then what he would wish to communicate to her is his *conditional* intention to kill her if she did not comply. Only if she believed him to have the conditional intention would she be rational in giving him her keys. Thus the court makes an invalid inference from, on the one hand, the unconditional intention to steal the car and the conditional intention to kill the driver if she did not comply to, on the other, the unconditional intention to kill the driver.

The court employs a (false) principle of the following form:

> (*) If S has a conditional intention to A if X and —— then S has an unconditional intention to A.

In fact, anyone who wishes to respond to the Mere Conditional Intent argument by claiming that under certain circumstances a person with a conditional intention has the forbidden unconditional intention must accept some principle of this form. In *Holloway*, the court fills the blank with something like "an unconditional intention to induce another to act so as to reduce the chances that S will A." What has just been argued is that the principle that the court employs is not true for the reasons necessary to support the court's argument. In fact, the principle the court employs is probably false, but the argument offered above need establish only something weaker; there are cases in which it is false, and the court fails to show that Holloway's is not such a case. Given burden of proof, this implies that the court cannot employ the principle in the way it does. But this does not imply that every nontrivial[13] principle of this form is not true in the way that it needs to be; maybe there is some way to fill the blank that results in a principle that can be used to support a court's attitude toward a particular defendant's employment of the Mere Conditional Intent argument. But

13. For our purposes, an instance of (*) is trivial just in case the value placed in the blank entails the consequent of (*) independently of the truth or falsity of the claim that the agent has a conditional intention to A if X. Such an account does not tell us the conditions under which a conditional intention is enough for mens rea since it identifies conditions sufficient for the unconditional intention even in the absence of the conditional intention.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

in fact there is not. To see why not, first consider as plausible-seeming an instance of (*) as one can imagine:

> (**) If S has a conditional intention to A if X and *S believes that X,* then S has an unconditional intention to A.

Interpreted as a psychological law of nature, this might be true. So interpreted, it might be true without exception, or it might be roughly true; it might be true of the vast majority of persons, actions, and conditions. After all, logically speaking, anything can cause anything else, so why should not a conditional intention and belief cause an unconditional intention? However, the plausibility of (**) does not derive from experience of the phenomenon of developing an unconditional intention when one has the relevant conditional intention and belief, much less from data of the sort examined by empirical psychologists that might (or might not) support (**). Rather, the principle appears plausible because it seems likely that in a rational person, the conditional intention and the belief that the condition obtains together play the same motivational role as an unconditional intention: Both motivate the agent to do A.[14] However, this point does not support the approach of someone who insists that some conditional intenders also have the forbidden unconditional intention. To have a conditional intention that causes you to act just as you would have acted had you had an unconditional intention is not the same as having the unconditional intention that the law forbids.[15] Someone who accepts that the law forbidding the crime requires unconditional intention—someone, that is, who accepts the first premise of the Mere Conditional Intent argument—needs something stronger than this. Such a person needs (**) itself, not the weaker claim that conditional intention and the relevant belief impose the same motivational pressure on rational agents as corresponding unconditional intentions.[16]

14. In fact, as Michael Bratman has pointed out to me, even this could be questioned. Imagine an agent who conditionally intends, say, to change his tire if it should be punctured in his drive across a bed of nails, believes that the tire will be punctured, but also intends to do everything in his power to prevent the puncture. This person's complex mental state plays a motivational role and a role in his practical reasoning (if he is rational) quite different from that of a person who unconditionally intends to change a tire. Still, if it is false that the relevant conditional intention and belief play the same motivational role as an unconditional intention, then efforts to respond to the Mere Conditional Intent argument by invoking something like (**) are weaker even than is suggested in the main text.

15. One might resist this conclusion on the grounds that if an agent does A intentionally, then the agent previously intended to A. This is what has become known in the action theory literature as "the simple view" of the relation between intentional action and prior intention. As attractive as the simple view is, it is also false. *See* MICHAEL BRATMAN, INTENTION, PLANS AND PRACTICAL REASON 111–138 (1987). For a valiant effort at defense of the simple view, *see* Hugh McCann, *Settled Objectives and Rational Constraints*, AM. PHIL. Q. 25–36 (Jan. 1991).

16. One might argue that because the conditional intention, when coupled with the belief that the condition obtains, plays the same motivational role as a corresponding unconditional intention, it is just as bad as the unconditional intention and is therefore sufficient for mens rea. This would be to adopt an approach of the same general sort as that discussed in Section II.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

Consider an example: A man points a gun at a driver, conditionally intending to kill her if she fails to give him her keys. She refuses, and so he comes to believe that the condition included in his conditional intention is satisfied; he kills her. After the formation of his belief and before his killing of the victim, did he come to have an unconditional intention to kill her? Perhaps, but nothing in the case as described supports the contention. The evidence supports only the claim that his conditional intention and belief motivated him to kill the victim. To insist that there was an intermediate between the conditional intention-belief pair and the action—namely an unconditional intention—is to invent a further mental state to play a causal role that can be played quite adequately by the conditional intention-belief pair.[17]

Perhaps there is reason to think that (**) is true on conceptual grounds. If so, then there is good reason to think that every defendant who has an unconditional intention and the belief that the condition obtains also has the relevant unconditional intention. But (**) is not true on conceptual grounds. Just as it is possible to have beliefs that are not closed under logical implication, it is possible to have sets of beliefs and intentions that are not. Arguably, although still not obviously, if we add a caveat to the effect that S is rational, then the principle is true on conceptual grounds. But there is no reason in general to assume that the defendant was in a rational state of mind at the crucial moments, much less in as rational a state of mind as he would need to be to have sets of beliefs and intentions that are closed under logical implication. Thus even this revised principle would not support the claim that defendants who admit to having had the relevant conditional intention and belief had the unconditional intention required for the crime.

These points can be extended to any nontrivial instance of (*). If the proposed principle that results from filling the blank in (*) with some value is intended as a causal claim, then it will inevitably involve identifying some set of states and events that, together with the conditional intention, play the same motivational role as a corresponding unconditional intention. But then any court that employs the proposed principle must offer additional evidence for the claim that the defendant has the relevant unconditional intention; the principle offered will not be enough. Alternatively, a principle constructed by filling the blank in (*) could help to resolve the problem posed by the Mere Conditional Intent argument if it were true on conceptual grounds without supplementation by a presumption of rationality on the

17. We might insist that even though there is no decisive evidence that the defendant with the conditional intention and belief had an unconditional intention, it is nonetheless likely that he did. This would be to insist that (**) is usually true or true of most people and most actions and conditions. However, notice that nothing in the argument just offered against a court's appealing to (**) turns on (**)'s being thought true without exception. So the same argument can be offered against someone who claims (**) only to be likely enough to support the claim that various defendants have the relevant unconditional intention.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

part of the agent. But what are the prospects for discovering a principle of the form of (*) that satisfies this requirement? In general, the fact that an agent has an intention to do one thing does not tell us that he also has an intention do something else; it rarely even provides *prima facie* support for such a contention. Why should the situation be any different when the two intentions are a conditional intention and corresponding unconditional intention? To assert that a principle of the form of (*) is true on conceptual grounds independently of presumptions of rationality is to assume falsely that there are necessary linkages between mental states. However, the only glue connecting one mental state to another is rationality; in the absence of rationality, almost any set of states can be found in a particular defendant's mind. Either way, then, a principle of the form of (*) fails to resolve the problem posed by the Mere Conditional Intent argument.

II.

The failure of efforts to saddle some conditional intenders with a corresponding unconditional intention encourages an alternative approach. Perhaps conditional intenders, despite lacking the explicitly forbidden unconditional intention, nonetheless have just as bad a mental state as someone with an unconditional intention. Those who adopt this approach, then, would say that what the relevant laws really ban are unconditional intentions *and other intentions that are just as bad.* This thought animates the approach offered in the Model Penal Code. The relevant section of the Model Penal Code reads as follows:

> When a particular purpose is an element of an offense, the element is established although such purpose is conditional, unless the condition negatives the harm or evil sought to be prevented by the law defining the offense. (Model Penal Code, §2.02(6))

The view is elaborated in the commentary:

> [I]t is no less a burglary if the defendant's purpose was to steal only if no one was at home or if he found the object he sought. The condition does not negative the evil that the law defining burglary is designed to control, irrespective of whether the condition is fulfilled or fails. But it would not be an assault with the intent to rape, if the defendant's purpose was to accomplish the sexual relation only if the mature victim consented; the condition negatives the evil with which the law has been framed to deal. (Model Penal Code, commentary on §2.02, p. 247)

Since there is some ambiguity in the Model Penal Code's (MPC) test as written, it is worth formalizing the proposed test. Here S is an agent; C is a type of act the performance of which requires intent to perform act A;

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

L is a law prohibiting performance of acts of type C; S's conduct is such that it would be an instance of C if S had the relevant intent; and S has the conditional intention to A if some condition X obtains:

> *The MPC Test*: S's act is not an instance of C *if and only if* If X obtains, then the evil that L is intended to prevent would not occur.[18, 19]

The motivating idea here is related to the motivating idea behind one aspect of the M'Naughten rule's test for legal insanity: If a person has delusional and false beliefs that influence his action, he is excused from legal responsibility only if it is the case that were those beliefs true, what he did would not have been illegal. The man who has the delusional belief that the person he killed was wearing a red shirt, when the victim's shirt was actually white, is not thereby excused, because even if the victim had been wearing a red shirt, killing him would not have been acceptable. Similarly, the thought goes, if the action the agent would have performed, had the condition governing his conditional intention been realized, is illegal, then the fact that his intention was merely conditional provides no excuse for his actual conduct.

The two questions to ask about the MPC test are: (1) Does it correctly identify the principled grounds for distinguishing between defendants with conditional intentions? and (2) Is it extensionally adequate? That is, does it draw the lines in the normatively correct way? I'll argue that the answer to both questions is "no." However, there is nonetheless an important lesson to learn from the MPC test.

To see a problem with the principled grounds that the MPC test offers for distinguishing between conditional intenders, put aside conditional intention for a moment and consider the following general proposal. Here all the variables are as in the formulation of the MPC test above, except that S is assumed not to have a conditional intention to A if X but instead an unconditional intention to B, where B is something distinct from A:

18. Cartwright, *supra* note 10, at 247–248, offers a test that is functionally equivalent to this test. Instead of considering whether or not the evil the law is intended to prevent would occur if the condition were satisfied, Cartwright confines his test to the question of whether or not the mens rea conditions of the crime are satisfied if the condition obtains. As will become clear from the discussion below, Cartwright's test succeeds no better than the Model Penal Code's.

19. In conversation, Herbert Morris has suggested to me that the framers of the Model Penal Code do not have this formal test in mind, despite the fact that the courts have treated the Code as though it were proposing this test. In the Model Penal Code's discussion of attempted crimes, for instance, it is never suggested that an agent with a merely conditional intention like Holloway's could be said to be engaging in an attempted crime. This implies that the Code is inconsistent in its treatment of cases of this sort. So, suggests Morris, the Code's remarks about conditional intention should be given much less emphasis in the interpretation of the Code than the court has given them. Although Morris is probably right about the intentions of the framers of the Model Penal Code—they probably did not intend that so many conditional intentions be thought sufficient for purpose—there is nothing else in the Code that can be used to determine which conditional intentions are and are not sufficient, and so it is not so unreasonable for the courts to have extracted this test from the Model Penal Code.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

(***) S's act is an instance of C *if and only if* If S were to do as he intends (that is, B), then the evil that L is intended to prevent would occur.

According to (***), a defendant's intention is as bad as that explicitly forbidden by the law in question, just in case acting on it would bring about the evil the law is intended to prevent. (***) yields unacceptable consequences. In fact, it does so whenever the purpose of the law is not to prevent just those acts that are intended by a defendant who has committed the crime the law bans. In fact, if the Supreme Court is to be believed, this is true of the federal carjacking law. In *Holloway*, Justice Stevens makes the following remark:

> Of course, in this case the condition that the driver surrender the car was the precise evil Congress wanted to prevent. (Majority Opinion in *Holloway*, n. 11.)

If this is true, then (***) tells us that a defendant should be held guilty of carjacking when he has no intention, conditional or otherwise, to kill but unconditionally intends to wrestle the keys from the driver's hand. After all, were he to succeed in doing as he intended, "the precise evil Congress wanted to prevent" would come to pass; it follows from (***) that such an intention is sufficient for the *mens rea* of carjacking, and this is an absurd result. The point is not merely, however, that (***) draws lines in a normatively unacceptable way. The point is that it does so because it identifies the wrong principled basis for the normative assessment of an intention. Our laws ban intentions, that is, for reasons other than the fact that were they acted upon, evils we wish to prevent would come to pass.

Nothing said so far shows there to be a problem with the MPC test. The MPC test, after all, is not the same as (***). It is, in effect, (***) restricted to those cases in which B is "to A if X," and it is quite possible, given only what has been said so far, that the problem with (***) just noted is avoided in this restricted set of cases. But is it? No. Imagine a pair of defendants both of whom threaten a driver, both of whom thereby induce the driver's compliance, and both of whom have a conditional intention. Imagine, however, that the first's conditional intention is to wrestle the keys from the driver's hand if she does not comply, and the second's is to kill her if she does not comply. According to the principles of evaluation of intention proposed by the MPC test, the intentions are equivalently bad, for *were the condition to obtain the evil the law is intended to prevent would come to pass.*

However, to say that the principles of evaluation that the MPC test proposes fail to draw a line between these two defendants is not to say that the test itself fails to draw a line. Quite the contrary; the test is to be employed in this case only to measure the value of the conditional intentions of those who conditionally intend *to kill* and not of those who conditionally intend

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

to take keys. The MPC test, that is, offers an account of the bar over which a conditional intention must go if it is to be as bad as a particular unconditional intention, but it stipulates that the conditional intention of the first of our two defendants is simply not to be judged by this standard. But what justifies this stipulation?

It is admitted that neither of our defendants has the same intention as that of the unconditional intender. So the difference between them must be evaluative: The conditional intention to kill must be worse than the conditional intention to take the keys, and this despite the fact that it is not worse *by the standard the MPC test supplies* (assuming that the "evil sought to be prevented by the law" is as represented by the majority in *Holloway*.)[20] It is true that one of the conditional intentions includes the action of killing in its content, and the other does not. Can that justify the stipulation? Perhaps someone who includes killing anywhere in his plans while taking a car from a driver needs to justify himself—and the MPC test tells us when this justification succeeds for conditional intenders—while those who do not are never required to offer such a justification. Presumably, on this view, the requirement to justify himself imposed on the defendant who conditionally intends to kill, a requirement not imposed on the defendant who conditionally intends to forcibly take the keys, derives from the fact that it is intentions *to kill*, and not intentions to take keys, that the statute explicitly bans. Thus the statute at issue is setting two standards: By banning intentions to kill, it sets the standard by which to judge if a conditional intender needs to justify himself, and by aiming at the prevention of car theft, it sets the standard by which to judge if he has succeeded in doing so.

But what justifies requiring, on the one hand, those with conditional intentions to kill to justify themselves, but on the other, not making the same demand on those with conditional intentions to do other things, like take keys? How can we appeal at this point to anything other than what the statute aims to prevent? There is something that the statute is aiming to prevent by banning intentions to kill. But surely anyone whose intention has just as good a chance of contributing to the occurrence of that evil as an unconditional intention to kill does must justify himself. But then we are back where we began: If the condition included in the conditional intention of the defendant who merely conditionally intends to take the keys obtained, the evil that the statute is intended to prevent—and which it banned intentions to kill in order to prevent—would come to pass. Surely, then, such a conditional intender owes us as much of a justification as the defendant with the conditional intention to kill. It follows that the MPC test lacks principled grounds for excluding consideration of the question

20. Notice that it would not help to insist that the majority has misidentified the purpose of the law against carjacking. It is always possible to concoct pairs of intentions that will have the same effects—and so will both bring about the evil the law aims to prevent—despite having different and differently objectionable objects. Under the right circumstances, even an intention to take someone's keys can cause or be likely to cause that person's death.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

of whether or not the defendant who conditionally intends to take the keys has an intention as bad as the intention to kill that the statute explicitly bans.

Thus the MPC test makes the same mistake that (***) generally makes: It does not provide *principled grounds* for excusing some defendants plainly deserving of excuse; this is not to say that it does not excuse them—sometimes it does—but it does not provide good reasons for excusing them.[21]

Despite this problem, however, there is something right about the MPC test. It is partially extensionally adequate; it provides the right resolution of a certain class of cases in which the Mere Conditional Intent argument is employed. Although, as will be suggested in Section IV, the test provides the right resolution of these cases for the wrong reasons, it is nonetheless useful to identify with precision the class of cases that it resolves correctly. To do so, start by noting that in applying the MPC test, we encounter two distinct sorts of cases in which the counterfactual conditional on the right side of the biconditional is true. The first class is those cases in which every relevant possible world in which X obtains is one in which S does not perform an act of type C.[22] In cases of this sort, the obtaining of X entails that the agent's conduct fails to satisfy some necessary condition for performance of a token of C. If the agent intends to take a certain object if it is his, for instance, then every possible world in which the condition is satisfied is one in which he does not steal, for stealing requires that what is taken does not belong to the person taking it. The Model Penal Code's rape example falls into the same class. The lack of consent of the other party, assuming that party is an adult, is required for the act to count as rape, so any possible world in which the consent condition is satisfied is one in which no rape occurs.

In the second class of cases, in at least some relevant possible worlds in which X obtains, S performs an act of type C, but circumstances nonetheless preclude realization of the evil that the law is intended to prevent. If the link between performance of the relevant act and realization of the evil the law is intended to prevent is a necessary link rather than a merely contingent causal link, then this class will be empty. However, there is no reason to assume it will be. Very often a law prohibits conduct the performance of which does not realize the evil the law is intended to prevent. Most laws prohibiting attempted crimes are like this. The law against attempted murder is intended to prevent people from murdering other people; anyone

21. The argument just offered was sharpened a great deal in response to helpful comments by a referee for this journal.

22. Notice that it would be possible for the agent not to perform an act of type C in the relevant possible worlds and yet still, in those worlds, realize the evil the law was intended to prevent. For instance, imagine that P intends and takes steps to kill Q if R's efforts to do so fail. As it happens, R succeeds in killing Q. Is P guilty of attempted murder? In the possible world in which R fails to kill Q, P is not guilty of attempted murder for, we can imagine, he succeeds. Still, in the relevant possible worlds he realizes the evil the law against attempted murder is intended to prevent, namely, the killing of another person. It follows that under the Model Penal Code's test P is guilty of attempted murder. Perhaps that is the right result.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

who attempts a murder and fails is guilty of a crime under this law, but his conduct fails to realize the evil the law is intended to prevent. Not all similar examples involve prohibited failures to act in a reprehensible way. Laws against speeding, for instance, are intended to prevent accidents, but the vast majority of rightly convicted speeders do not cause any accident through their conduct. Thus if the evil the law is intended to prevent is merely a contingent causal consequence of some of the behavior the law bans, then there are possible worlds in which the condition is satisfied and the agent performs an act of type C but fails to realize the evil in question.

However, any case that falls into this second class is treated incorrectly by the MPC test. Presumably the framers of the Model Penal Code do not mean to imply that in general an agent is not guilty of a crime prohibited by a particular law when his action fails to realize the evil the law is intended to prevent. The framers of the Model Penal Code want to allow for criminal attempts and for the prohibition of acts that by luck fail to realize the harms for the sake of the prevention of which they are prohibited. But then the test excuses agents on the basis of the features of a counterfactual scenario that would not excuse the agent were it actual. The counterfactual scenario is the one that the agent is committed to realizing should the relevant condition obtain. Since in the counterfactual scenario the defendant takes steps to perform A, he commits an act of type C in the counterfactual scenario. It follows that the defendant is committed (conditionally, but still committed) to the realization of a state of affairs in which he performs a crime. How can such a commitment on his part excuse? For example, imagine that a person possesses a large amount of an illegal drug and conditionally intends to sell it to a particular buyer if the stock market goes down. But as it happens, if the stock market goes down the buyer will not have sufficient funds to buy the drugs. In possible worlds in which the stock market goes down, we can assume that the agent satisfies the conditions of possession with intent to distribute, but he does not succeed in distributing his drugs.[23] However, distribution of illegal drugs is the evil that laws against possession with intent are intended to prevent, and so the relevant evil does not occur if the relevant condition obtains. It follows that this is a case in which, under the Model Penal Code's test, the agent would not be guilty of possession with intent to distribute. The agent's conditional intent, under the MPC test, is insufficient for *mens rea*. But if the man who makes a failed effort to sell his drugs when the stock market goes down is guilty of possession with intent, why should the unrealized possibility of failure excuse the man who is lucky enough never to have to try to sell the drugs?

In the example just offered, the obtaining of the condition "negatives" the evil the law is intended to prevent by breaking the causal link between the defendant's behavior and the evil in question. In other cases, the

23. Any adequate semantics of counterfactuals will imply this result. That is, the result is not dependent, for instance, on any particular similarity metric for possible worlds.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

condition might "negative" the evil by necessarily precluding its occurrence, but without precluding performance of an act of type C. For instance, let us assume that the purpose of a law against possession of marijuana with the intent to distribute is to prevent unhealthy consumption. Under the MPC test, the Mere Conditional Intent argument offered by a defendant who conditionally intends to sell marijuana to a particular buyer if consumption of it will be healthy for that buyer should be honored; after all, if the condition were satisfied, the evil the law is intended to prevent would necessarily fail to come to pass. But as in the previous example, the MPC test then excuses a defendant who is conditionally committed to a course of conduct that would be criminal were it actual.

So all of the cases that the MPC test resolves correctly fall into the first class: They are all cases in which the obtaining of the condition precludes the possibility of performing an act of the type of the crime in question. However, the MPC test does not resolve all of these cases correctly.[24] For instance, say that a person kills his uncle with the conditional intention to kill him if he can legally inherit the uncle's property. Further, imagine that he grabs what he believes will be his only opportunity to kill the uncle, despite the fact that he does not know the answer to the legal question. He is charged with murder. In any situation in which the condition is satisfied, the killing of the uncle will not be murder—after all, a murderer cannot legally inherit the victim's property. Thus if a conditional intention generally failed to be sufficient for *mens rea* if the condition precludes the satisfaction of some necessary condition for the offense, then it would follow that this man is not guilty of murder but of some lesser offense that does not require the intent to kill. And this is an absurd result.[25]

24. The Model Penal Code will not provide an adequate resolution of any case in which the obtaining of condition X precludes the possibility that S performs an act of type C by precluding the possibility that S satisfies C's intent requirement. Imagine, for instance, that an agent conditionally intends to A if he does not change his mind. Thus any possible world in which the condition obtains—the agent does not change his mind—is one in which he maintains his conditional intention; if he were to have an unconditional intention in such a possible world, he would have changed his mind, contrary to hypothesis. If such a person fails to perform an act of type C in such a world, it must be because C's intent requirement is not satisfied by S's conditional intention. Of course, that is not something that we can learn from the MPC test, since the MPC test directs us to answer the question of whether or not S's conditional intention is sufficient by examining a possible world in which he possesses that very conditional intention. In short, in cases of this kind we encounter circularity in the application of the MPC test. We do not encounter this circularity problem if the obtaining of condition X thereby falsifies some condition required for performance of an act of type C. This is the case in our earlier examples in which the satisfaction of the condition precludes performance of an act of the relevant type. The satisfaction of the condition involved in the conditional intention of the man who intends to take the object if it is his falsifies a condition on stealing that is independent of the intent required for stealing. We do not have to determine whether or not the mere fact that the man's intention is conditional, independently of the content of the condition, precludes satisfaction of the intent requirement of stealing. From here on, it is assumed that conditions that preclude satisfaction of a necessary condition for performance of the crime are all intent-independent.

25. We might think that what is special about this example is that the connection between the falsity of the condition (its not being legal to inherit) and the type of act in question

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

This way of analyzing the example just offered might be resisted. First, someone anxious to defend the MPC test might insist that the defendant in this imagined case actually has the unconditional intention to kill when he kills his uncle. Notice, however, that this response is not available at this stage in the argument. It amounts to insisting that something went wrong in the argument in Section I for the claim that there can be no grounds for thinking that a conditional intention, when conjoined with any set of circumstances, simply amounts to or entails the possession of an unconditional intention; the fact that the nephew kills his uncle does not imply that his conduct was guided by anything more than a conditional intention to do so. Retreating from this position, then, it might be argued that the nephew possesses the *mens rea* of murder despite the fact that he acts on a merely conditional intention to kill and despite the fact that his conditional intention is not, by the MPC test's standards, as bad as an unconditional intention. Perhaps his conditional intention satisfies *mens rea* because he *aims* at the uncle's murder even though he does not unconditionally intend it; perhaps it is enough, that is, that the murder was what the Model Penal Code calls his "conscious object" [see Model Penal Code Section 2.02(2)(a)].[26] However, this suggestion does not help to defend the MPC test. The problem is that your "conscious object" when you have an intention is just that which you intend, or the content of your intention. The very problem we are trying to solve is that one's conscious object is sometimes conditional, as in this case. And so this move just sends the defender of the MPC test back to the beginning: Under what conditions is it just as bad for your conscious object to be conditional as unconditional? The result is that the Model Penal Code does not provide the resources to explain why the mental state of the murderous nephew is just as bad as that of the unconditional intender.

The lesson to draw is this: The MPC test provides the right resolution of most cases in which the condition governing the defendant's intention precludes satisfaction of a necessary condition for the crime. What will

(murder) is not constitutive. (Thanks to Rob Kar for pressing this point.) While it is in fact necessary for an act to be murder that it is not legal to inherit from the victim, it is not part of the definition of murder that this be so; the fact has its basis instead in precedents that take the act of murder to be independently defined. Perhaps, then, the range of cases in which the MPC test gives the right verdict should be identified as follows: all those in which satisfaction of the condition precludes satisfaction of some necessary and constitutive condition for the crime in question. This, however, is not right. We can imagine a legal system in which it is part of the definition of murder that the killer cannot inherit the victim's property. Under such a system, a defendant could show himself to be not guilty of murder by showing that he was to inherit from the victim. Such a system would be morally perverse, but would it be a system in which our killer's conditional intention was insufficient for mens rea? Perhaps, but not clearly. In our example, the killer does not know the answer to the legal question of whether he can inherit when he kills his uncle; he may not even know that his killing will fail to count as murder if the condition is satisfied. Why should his mental state be thought to be less bad than that of the man who commits the same acts from the same motives, intentions, and beliefs under our legal system? More about this issue *infra* note 32.

26. Thanks to a referee for this journal for pushing this objection.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

emerge in Section IV is that the account to be offered of the conditions under which a conditional intention is and is not sufficient for *mens rea* will explain, first, why the MPC test gives the right results in most of these cases, and second, why it nonetheless gives the wrong result in some of them. Once we see the real conditions under which a conditional intention is and is not sufficient for *mens rea*, we are in position to see why the MPC test sometimes succeeds and sometimes fails to track the fundamental features of guilty and innocent conditional intentions.

The MPC test focuses entirely on the *content* of the conditional intention and provides an account of the circumstances under which an act that falls under the description "to A if X" is innocent in comparison to other acts that fall under the description "to A." As has been argued in this section, the MPC test succeeds when applied to only a very limited domain of cases in which the Mere Conditional Intent argument is employed and fails for a much larger class of cases, including cases like *Holloway*, in which it has guided the court's decision. The MPC test, however, does not consider what sort of commitment *to A* is manifested in a conditional intention to A if X. Rather, it considers "to A if X" to pick out a type of action distinct from that picked out by "to A" and then proposes conditions under which actions of the former sort are not illegal by the standards of the law banning the crime in question (namely, C). Thus the MPC test overlooks the possibility that an agent could be excused from responsibility on the grounds that his intention is merely conditional not because he is thereby committed to something different from what the law bans him from being committed to, but because his commitment to what the law bans (namely, A) is not as robust as that of the person who is unconditionally committed to that same action. Conversely, the MPC test overlooks the possibility that an agent with a conditional intention could be guilty not because he is committed to something different from A that is just as bad, but because his conditional intention amounts to a commitment *to A* that is sufficient to warrant calling his conditional intention criminal. Thus if we are to have a satisfactory account of the conditions under which a conditional intention is and is not sufficient for *mens rea*, we need to know how the commitment to A of someone with a conditional intention to A if X differs from the commitment to A of someone with an unconditional intention to A. It is to that issue that we now turn.

III.

A great deal is already known about the commitment involved in unconditional intentions and how it differs from, say, the sort of commitment to a course of conduct of a person who desires to engage in that conduct. There is a closely related and large literature on the question of what intentions are and whether or not they are reducible to complexes of beliefs and/or

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

desires.[27] The question with which that literature is occupied, sometimes explicitly and sometimes less so, is whether or not complexes of beliefs and desires are capable of constituting the sort of commitment to conduct that is distinctive of intention. Of particular importance to the question of the kind of commitment involved in conditional intentions, however, is a broadly schematic point that has been more or less accepted by action theorists since the publication of Michael Bratman's *Intention, Plans and Practical Reason*:[28] Intentions, unlike desires, are rationally stable; intentions, that is, involve a certain kind of commitment to what is intended that is constituted by the fact that agents who intend while failing to be committed to what they intend are guilty of some form of irrationality. If we know the conditions under which an intending agent responds to his intention rationally or irrationally, then we know what it is to be committed in the way that is distinctive of intention.

Of particular importance for our purposes is the emergence of the rational stability of intention in the influence that intentions have on rational deliberation after the time of intention formation. (Also important are the rational pressures placed on an agent's actions at the time of intention execution, but we will ignore such pressures here.) Of particular importance to an agent's moral and legal responsibility is the agent's pattern of response to reasons, particularly when that pattern is self-imposed. Whether we take criminal punishment to be justified for reasons of deterrence, rehabilitation, or retribution, the defendant's pattern of response to reasons is important to his legal responsibility. Those whose conduct is expressive of a settled pattern of responsiveness to reasons are harder to deter, harder to rehabilitate, and guilty of more morally objectionably conduct than those whose conduct is not guided, or is guided less robustly, by rationally structured deliberation.[29] Thus the structure imposed on deliberation by intentions is part of what makes intent relevant to moral and legal responsibility. We can see that intentions impose structure on further deliberations by considering commonplace examples: A rational person who intends this morning to buy milk this afternoon does not deliberate this morning about how to go about getting milk this evening; it would be irrational to do so. Nor does he deliberate about whether to do things this morning that would (relative to his beliefs) preclude the possibility of his getting milk this afternoon. He

27. For a start, *see* Donald Davidson, *Actions, Reasons and Causes, in* Essays on Actions and Events 3–19 (1980); Michael Bratman, *Davidson's Theory of Intention, in* Faces of Intention 209–224 (1999); Wayne Davis, *A Causal Theory of Intending*, 21 Am. Phil. Q. 43–54 (1984); J. David Velleman, Practical Reflection 121–126 (1989).

28. Bratman, *supra* note 15. It is a vast understatement to say that my general remarks here regarding the nature of the commitment involved in intending and the particular rational pressures under which an intending agent labors are influenced by Bratman's views. They are meant, rather, to be merely restating points that Bratman has made.

29. John Fischer and Mark Ravizza have developed this sort of position. *Cf.* John Fischer, The Metaphysics of Free Will (1994); John Fischer & Mark Ravizza, Responsibility and Control: A Theory of Moral Responsibility (1998).

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

deliberates about what to do tonight under the assumption that he will have already gotten milk this afternoon, and he deliberates about what to do this morning under the assumption that he will be getting milk this afternoon.[30]

To understand better the nature of such pressures, it helps to have a rough model of deliberation in hand. Deliberation, for our purposes, is a norm-governed psychological process. It consists of considering scenarios or situations, where these are understood to be states of the world at particular times, and determining, for each considered scenario, what course of conduct to undertake by weighing reasons for and against each possible action that might be performed in the scenario under consideration. To say that this psychological process is "norm-governed" is to say that it can go well or badly; the agent can engage in it in a rationally acceptable or unacceptable manner.

The scenarios that one considers in determining what to do are not total states of the world. What is and is not included in a particular scenario are a function of what the agent flat-out believes about what the state of the world is or will be at a particular time of action. They include all and only those features of the world that the agent accepts as being in place at the time of action. Since there are many possible features of the world that we neither flat-out believe to be in place nor flat-out believe *not* to be in place, scenarios neither include nor exclude such features. If I am deliberating about what move to make to open a chess game, the standard opening positions of my and my opponent's pieces are included in the scenario I consider. Excluded, however, are the positions of the pieces after the second move. This exclusion is not a result of the futurity, relative to the time of action, of the positions of the pieces after the second move. Scenarios typically include features of the world in the past and future relative to the time of action, since agents typically flat-out believe various past and future features to be in place. Rather, the positions of the pieces after the second move are not included in the scenario I am considering when deciding how to open the game because I do not *flat-out* believe that my opponent will move this way or that in response to my opening moves. Still, I might know my opponent very well and attach a high probability

30. Notice that this point is highly schematic in the following sense: Two parties could agree that intentions place rational constraints on deliberation and disagree about what those rationality conditions are. Take an example that has been hotly disputed by action theorists: We might disagree about whether or not a rational agent who intends to act will be guided in his deliberations by a belief that he will act or merely by the absence of a belief that he will not. Cf. H.P. Grice, *Intention and Uncertainty*, 57 PROC. BRIT. ACAD. 263–279 (1971); Donald Davidson, *Intending*, *in* ESSAYS ON ACTIONS AND EVENTS 83–102 (1980); Gilbert Harman, *Practical Reasoning*, 29 REV. METAPHYSICS 431–463 (1976); BRATMAN, *supra* note 15, at 37–41; VELLEMAN, *supra* note 27, at 113–121. The point made so far cuts across disagreements of this sort: If an agent intends to do something, then there exist norms of rationality governing his deliberation that are applicable to him solely by virtue of the fact that he has an intention. To know the relevant norms of rationality is to know what it is to be committed to conduct in the manner involved in intention. A theory of conditional intention, then, will specify the norms of rationality governing the deliberation of an agent with a conditional intention.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

to his moving in a particular way in response to a certain opening move. This probabilistic belief can influence what it is rational for me to choose as my opening move—perhaps I know that there is a high probability that my opponent will make an error when confronted with a particular opening move—but the error is not included in the scenario that I am considering, since as likely as I think it is that he will make the error, I do not flat-out believe that he will. The idea of a "scenario," that is, is the idea of what is *given*, or for practical purposes immutable. This conception of a scenario will turn out to be of some importance in what follows.

There are at least two sorts of rational pressures regarding deliberation under which our intentions place us: There are those regarding what courses of conduct we are *not* to consider pursuing when reflecting on a particular scenario, and there are those regarding what sorts of conclusions we must reach regarding what to do in the scenarios that we consider. Bratman captures the first sort of pressure by saying that intentions "filter options" for consideration in deliberation: If you intend to A, then you must not consider courses of conduct that are incompatible with your performing A. This does not imply that you consider only courses of conduct in which you do A; some courses of conduct are neutral with respect to the question of whether or not you A. To violate this demand is to waste deliberative effort. One of the benefits that we gain from the formation of intention is the simplification of our further deliberations; if you intend to do something, then there are fewer courses of conduct to consider in future deliberations, for you need not (in fact, ought not to) consider courses of conduct incompatible with what you already intend.

A standard example of the second sort of pressure—pressure regarding what conclusions to reach in deliberation—derives from the need for means-end coherence in one's intentions; if you intend to A, then with respect to every scenario that you consider, you must reach intentions to take the means of A-ing in that scenario, and you must reach intentions to perform acts that are required in addition to A in order to reach the end that A serves.

What sorts of pressures of the second sort—pressures with respect to what conclusions to reach in deliberation—are imposed on an agent with a conditional intention? It is important to note that an agent who conditionally intends to A if X is under rational pressure to adopt the necessary means of A-ing in many scenarios that he considers. So, to take a typical example, say that I am building a house and quite rationally believe that it will be possible to dig the holes for the foundation in twenty man-hours; I set aside the necessary funds. However, I conditionally intend to dig for thirty man-hours if twenty is not enough. My conditional intention to dig for thirty man-hours is a contingency plan; it is what I plan to do if my primary plans fall through. Now, if I begin digging before I have the money to pay for the extra ten man-hours and I have no reason to expect that I will get the money in time should the extra time be needed, then I am acting irrationally. That is,

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

conditional intentions often place agents under rational pressure to adopt the means to the act that they merely conditionally intend to engage in. Given my conditional intention in this example, I am committed to the act of digging for thirty man-hours at least to the degree to which a person who is committed to constraining his deliberations by the need to take necessary means is so committed.

In fact agents with conditional intentions are under pressure to adopt means to their conditionally intended acts even if they attach a very low probability to the obtaining of the condition, as long as they do not flat-out believe that the condition does not obtain in the scenario under consideration. Say that you owe me $10,000 and if you pay me by Friday, I can use the money to buy a boat that will be auctioned on Saturday. So I conditionally intend to buy the boat if you pay me by Friday. Further, imagine that in order to participate in the auction, I have to register for it on Monday. Now I am deliberating about what to do on Monday. The scenario that I am considering neither excludes nor includes your paying me by Friday; I do not flat-out believe that you will or that you will not. Now, if we assume that there is no cost to registering—there is no fee involved, it takes no time whatsoever, it does not keep me from doing anything else that I need to do—then I am under rational pressure to decide to register for the auction on Monday. I am under such pressure as long as there is any chance at all that you will pay me by Friday. Imagine, even, that your paying me by Friday is contingent on your winning the lottery on Thursday. Still, as long as I do not flat-out believe that you will not win, and as long as there is really no cost whatsoever to registering for the auction, I am under rational pressure to register.

Of course, this pressure can be overridden, and whether or not it is overridden is a function in part of the probability that I attach to the obtaining of the condition—your paying me by Friday. Say, for instance, that it costs $10 to register for the auction, and I think the probability of your paying me by Friday is one in 10,000. In this case, the value in dollars of registering for the auction is just the expected benefit of doing so, minus the cost, and the *expected* benefit is a function of the probability I attach to your paying me; that is, the value of registering is $[(\$10{,}000 * .00001) - \$10] = -\$9$. That is, it is not worth registering. Still, even in this case, my conditional intention places me under rational pressure to adopt the means of buying the boat; it is just that I am under stronger rational pressure not to decide on any action from which I can expect a loss. In fact, there is nothing special about conditional intentions in this regard. If I *un*conditionally intend to attend the auction, I am under rational pressure to register, but this pressure would be overridden if the expected value to me of registering is not positive.

What of the other sorts of rational pressure on deliberation placed by intentions? What sorts of pressures with respect to what not to consider are placed on an agent by his conditional intentions? Notable in this regard is less what pressures are imposed by conditional intentions than what

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

pressures are not imposed. In particular, *a person who has a conditional intention is often free to consider courses of conduct incompatible with the conditionally intended action*. If I conditionally intend to dig for thirty man-hours if twenty is not enough, I am nonetheless perfectly free to consider various ways of using the money for the extra man-hours that are incompatible with using it to pay for extra digging. I would be irrational were I to take steps to use that money in some way that precludes the possibility of using it for the extra man-hours, but I am still not guilty of any irrationality if I *consider* such courses of conduct. After all, the stricture against considering acts incompatible with what I intend comes from the need to avoid wasting deliberative effort.

But when I neither flat-out believe that I will need to use the money for the extra man-hours of digging nor flat-out believe that I will not, I am considering what to do in a scenario in which I very well may not use the money to pay for the extra man-hours of digging. So I do not waste deliberative effort when considering alternative ways to use the money in such a scenario. This, then, is an important difference between a conditional and unconditional intention: An unconditional intention to dig for thirty man-hours places me under pressure not to consider ways of using the money to do something incompatible with paying the diggers—to consider such actions would be a waste of my time—but a conditional intention often places no such pressure on deliberation.

These results can be extended. Referring to the table below, compare the rational pressures faced by an agent who unconditionally intends to A (call him "S1"), and an agent who conditionally intends to A if X (call him "S2").[31] There are three sorts of scenarios that could enter into either agent's deliberations: "X-scenarios"—those in which X obtains; "not-X-scenarios"—those in which X does not obtain; and "X-neutral scenarios"—those that are neutral with respect to the question of whether X obtains. There is no difference between S1's and S2's deliberations about what to do in X-scenarios (see boxes a1 and b1); both are under pressure not to consider courses of conduct that exclude A-ing and are under pressure to reach conclusions about what to do that are means-end coherent with A. There is a great deal of difference between S1's and S2's deliberations with respect to not-X-scenarios (boxes a2 and b2): S1 labors under the same pressures as he does when considering what to do in X-scenarios, while S2's

31. Both Michael Bratman and Rob Kar have noted in conversation that I am assuming that S1's unconditional intention is not such that S1 would abandon it if X were to obtain. A further question, therefore, remains about the relation between the rational pressures placed on us by intentions that are subject to conditions and those placed on us by, on the one hand, corresponding conditional intentions, and on the other, unconditional intentions that are not subject to the same conditions. The discussion here, however, is limited to conditional intention proper and does not extend neatly to intentions that are subject to conditions that are not properly included in their content. Finnis, *supra* note 10, at 165–170, argues that there is no difference in culpability between having a conditional intention, on the one hand, and an unconditional intention subject to the same conditions, on the other.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

conditional intention places him under no pressures whatsoever either with respect to what not to consider or with respect to what conclusions to reach. With respect to their deliberations about what to do in X-neutral-scenarios, however, S1 and S2 labor under the same pressures with respect to what conclusions to reach (box b3) but do not labor under the same pressures with respect to what to consider (box a3). Where S1 can consider only courses of conduct that do not exclude him A-ing, S2 is free to consider courses of conduct that do exclude that possibility as long as those courses of conduct are compatible with X failing to obtain. However, both S1 and S2 are required to adopt courses of conduct that are means-end coherent with their A-ing. This is obviously the case for S1, but it is also true for S2; when considering what to do in scenarios that do not specifically exclude the possibility that X obtains, S2 is under pressure to adopt means of A-ing and otherwise to form only plans that are means-end coherent with his A-ing.

		a	b
		Is the agent under pressure to consider only courses of conduct that do not exclude him A-ing?	Is the agent under pressure to adopt courses of conduct that are means-end coherent with his A-ing?
1	X-scenarios	S1: Yes S2: Yes	S1: Yes S2: Yes
2	Not-X-scenarios	S1: Yes S2: No	S1: Yes S2: No
3	X-neutral-scenarios	S1: Yes S2: No	S1: Yes S2: Yes

To illustrate, consider an agent who intends to steal a car the following morning by inducing the driver to hand over her keys and he unconditionally intends to kill her. In all of his deliberations about what to do between now and the time of the crime and about what to do during and after, the agent is consistently under pressure from his unconditional intention to consider only courses of conduct that include him killing the driver and to select only courses of conduct that are means-end coherent with his performance of the murder. In deciding what to do this evening, for instance, he might be under rational pressure to elect to buy ammunition before the stores close; in deciding what to do tomorrow evening, he might be under rational pressure to elect to take a boat trip so as to dump the body.

Now consider Holloway's deliberations the day before his crime. He intends to steal a car tomorrow morning and conditionally intends to kill the

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

driver if she does not comply with his demand to hand over her keys. In deliberating about what to do if the driver does not comply (X-scenarios), Holloway is under pressure to consider only courses of conduct that include him killing her. Further, he is under pressure to plan to do various other things that are necessary means to that end: He needs to plan to bring a loaded gun, say, and to plan to wash the car's front seat, perhaps, before selling it. In deliberating about what to do if the victim does comply (not-X-scenarios), Holloway, however, is unconstrained by the fact that he has a conditional intention to kill her if she does not. As far as the demands imposed by his conditional intention go, he is free to consider any courses of conduct that he likes and to adopt whichever he pleases.

Where he is in a somewhat strange position is with respect to his deliberations about what to do in situations in which it is unestablished whether or not the victim complies. So, for instance, when considering what to do this evening, his conditional intention does not preclude him from considering going on a bender that would leave him sick in bed for the following day and thus unable to perform the robbery, while the agent with the unconditional intention is precluded from considering this option. (It is true that Holloway's unconditional intention to steal a car tomorrow morning places him under independent pressure not to consider going on a bender, but we are confining ourselves to consideration of the pressures placed on him by his conditional intention to kill.) Still, at the same time as he is free to consider courses of conduct that would leave him sick in bed the next day, he is under pressure to adopt the means of killing the victim the next day. His conditional intention might, for instance, place him under pressure to plan to take a trip to the gun store, since he is out of ammunition.

For simplicity in exposition, let us refer to the part of a person's deliberation that responds to the pressures indicated in, say, box a1 in the above table as his "a1-deliberations," those that respond to the pressures indicated in b2 as his "b2-deliberations," and so on; each of these sorts of deliberation are sections of the norm-governed psychological process of deliberating, sections that are identified by appeal to the particular norms governing them. Deliberations of these various sorts can be grouped in illuminating ways. For our purposes, of particular interest is the contrast between a2-, a3-, and b2-deliberations, on the one hand, and a1-, b1-, and b3-deliberations, on the other. In the former group, differences in the nature of the commitment to the relevant course of conduct are found between agents with conditional and unconditional intentions, while in the latter group similarities are found. Let us call the former group "distinct-deliberations," and the latter, "similar-deliberations." Those who think that a particular defendant's employment of the Mere Conditional Intent argument should be honored think that with respect to that defendant, what is salient are his distinct-deliberations; they think, that is, that his conditional intention is indicative of deliberations structured significantly differently from those of the unconditional intender. On the other hand, those who think that a particular

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

defendant's employment of the Mere Conditional Intent argument should be rejected think that with respect to that defendant, what is salient are his similar-deliberations. The next section proposes a test for determining in particular cases whether distinct- or similar-deliberations are salient, and thus provides a test for determining when the Mere Conditional Intent argument should be honored and when it should be rejected.

IV.

Let us return for a moment to the Model Penal Code's test for determining if a conditional intention is sufficient for *mens rea*. Recall that in Section II we reached the conclusion that, barring certain problematic cases, the Model Penal Code's test does correctly identify a set of conditional intentions that are insufficient for *mens rea*: those in which the obtaining of X precludes the agent's satisfaction of some necessary condition for performance of an act of the sort of the crime in question. We are now in a position to see why this kind of conditional intention is usually insufficient for *mens rea*: An agent who has such a conditional intention and is deliberating about what to do in X-scenarios cannot be thereby deliberating about whether to commit the crime in question; after all, every X-scenario is one that precludes the possibility of committing the crime. So in such cases, the fact that there is perfect similarity between the pressures faced by S1 and S2 in their deliberations about X-scenarios (in deliberating about what to do in X-scenarios, both engage exclusively in similar-deliberations) is not salient. That aspect of both S1's and S2's deliberations is perfectly innocent, and thus the case for thinking the defendant to have met *mens rea* cannot be made by appeal to the role of his conditional intention in structuring his deliberations about X-scenarios.

However, we are also now in a position to see what is problematic about the problematic cases for the claim that it is *generally* the case that a conditional intention is insufficient for *mens rea* when satisfaction of the condition precludes satisfaction of some condition for the offense. Recall our earlier example: A man kills his uncle with the conditional intention to kill him if he can legally inherit his estate, and he has no idea when he does the killing whether or not he can legally inherit.[32] However, he thinks that the

32. Recall the earlier suggestion about this example; perhaps the MPC test gives the right verdict only if satisfaction of the condition is constitutive of the nonperformance of the crime in question. (An objection to this suggestion was offered *supra* note 25.) We are now in position to see why this suggested position is incorrect: What matters to culpability is the deliberation motivating the agent's action. But the question of whether the statement "Murderers cannot inherit their victim's property" is true by virtue of the definition of murder or true by virtue of precedent is irrelevant to the structure of the agent's deliberation. What matters for practical purposes is that it is true, not why it is true. This is why it is a mistake to distinguish between the culpability of the person under our legal system who conditionally intends to kill his uncle and the person with the same conditional intention but who is prosecuted under a legal system in which the illegality of inheritance is part of the definition of murder; the deliberations of these two people are structured no differently.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

time at which he kills him is the only chance that he will have to do so. It seems clear that his conditional intention is sufficient for the *mens rea* of murder, despite the fact that any situation in which he can legally inherit is one in which he did not commit murder. Notice what has occurred: The killer was guided in his conduct by the outcome of his deliberation with regard to an X-neutral-scenario. He was unsure if the situation in which he found himself was one in which it was or was not legal to inherit, but his conditional intention placed him under pressure to adopt the means of killing his uncle in such a scenario. That is, he was guided by deliberations in which he faced pressure to adopt means of the conditionally intended killing, and so he acted on a piece of deliberative reasoning that was identical to that of the person who unconditionally intends to kill. The MPC test gives us a reason to ignore certain deliberations in assessing *mens rea*: We can ignore the agent's deliberations about what to do in X-scenarios just in case the obtaining of X negates some necessary condition for the crime. But other deliberations remain that in this case motivate the agent's behavior. Since these deliberations are just like those of an unconditional intender, it follows that the defendant's conditional intention is sufficient for *mens rea*.

From this we can learn something about what a test for determining whether or not a conditional intention is sufficient for *mens rea* must look like: It must take a section of an agent's deliberations to be relevant to that assessment only if it does not exclusively involve reflection on scenarios that preclude the performance of the crime in question by precluding satisfaction of a necessary condition for the crime.

Further progress can be made in developing the test we are after by noting that, with good reason, the having of an intention by itself is never a crime in the United States. Even crimes of attempt require, in addition to the relevant intention, that the agent took substantial steps of some sort toward doing what he intended. Therefore the fact that an agent's intention places him under rational pressure to do something of which we disapprove—to adopt means to the performance of the intended act, for instance, or even just to deliberate to some further conclusions—does not by itself indicate that the agent is guilty of anything. The agent must also take some steps to comply with the rational pressures that his intention imposes. Thus a difference or similarity between the rational pressures faced by an agent with an unconditional intention and those faced by an agent with a conditional intention are of no relevance to the criminality of the agent's conduct except when that difference or similarity manifests itself in action of some kind. So in determining whether or not a conditional intention is sufficient for *mens rea*, we need to look at the deliberations that animate the conduct in which the agent engages, and we need to ignore deliberations that do not manifest themselves in conduct of any sort.

We are now in position to describe both a set of cases in which a conditional intention is not enough for *mens rea* and a set of cases in which it is. (A remaining set of cases will require additional resources.) To see this, first

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

reflect on the case of United States v. Rushlow.[33] Rushlow was a soldier who went absent without leave in order to work to support his poverty-stricken family. He was charged with the crime of being absent with the intent never to return, an offense holding a much more serious penalty than the offense of merely being absent. He argued in his defense that he intended to return if his brother came home and got a job sufficient to support the family. The court honors Rushlow's use of the Mere Conditional Intent argument, ruling that his conditional intention not to return if his brother never came home was not sufficient for the *mens rea* of the offense.[34] The case is a bit different from *Holloway* and *Irwin* in that in *Rushlow* the defendant cites his conditional intention *to return if his brother did come home and get a job*—rather than his conditional intention *not to return if the brother did not*—in his defense. Is there any significance to this? Sometimes in colloquial English, a conditional intention to A if X will be expressed with the words, "I intend not to A if not X." So understood, Rushlow's statement to the effect that he conditionally intended to return if his brother came home amounts to no more than a statement of the claim that his commitment to the forbidden act of never returning was merely conditional.

If this is all that Rushlow is saying, then his claim to have conditionally intended to return if his brother came home is just a statement of the Mere Conditional Intent argument, offered without further support for any of the argument's premises. Another possibility is that Rushlow is saying that he unconditionally intended to return but that he recognizes also that he is likely to have changed his mind should his brother not have come home. Understood in this way, Rushlow is not employing the Mere Conditional Intent argument at all; he is simply denying that he has any intention never to return, conditional or unconditional. In either of these two ways of understanding the case, then, the fact that Rushlow appeals to a conditional intention to return—rather than a conditional intention not to—is of no significance. However, there are two other ways to interpret Rushlow's statement.

First, Rushlow could be understood to be claiming that he did indeed have a conditional intention never to return but that he really never gave any weight in his deliberations to the possibility that his brother would never return—that the condition would be satisfied. He simply did not consider that eventuality or make plans about what to do should it occur. So understood, Rushlow's deliberations proceeded just as though he had a

33. 2 U.S.M.C.A. 641 (1953 CMA).

34. The Mere Conditional Intent argument arises in Rushlow with respect to jury instructions. The defense employs the Mere Conditional Intent argument to support its contention that the jury should be instructed that if they find Rushlow to have only conditionally intended never to return, that they should find him innocent of being absent without leave with intent never to return. The judge in the case honored the argument. However, the jury reached the conclusion that Rushlow actually unconditionally intended never to return, and so found him guilty.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

flat-out belief that his brother would return; so understood, he is claiming that the only sort of deliberation in which he engaged arose from reflection on not-X scenarios, since every scenario that entered into his deliberations was one in which not-X held. If such a contention is to believed, then none of Rushlow's actions were guided by similar-deliberations, since no similarities are to be found between the pressures faced by conditional and unconditional intenders in their deliberations about not-X scenarios. It follows that if Rushlow is to be believed, his conditional intention never to return if his brother never came home played a role in his practical reasoning that was entirely distinct from that of an unconditional intention never to return and so was insufficient for the *mens rea* of the crime.

However, Rushlow's claim to have conditionally intended to return if his brother came home (a conditional intention to not-A if not-X) could be understood in another way as well.[35] Consider the pressures encountered by a person with such a conditional intention when reflecting on not-X scenarios: Such a person is under pressures that push in exactly the opposite direction from those encountered by someone with an unconditional intention not to return (an unconditional intention to A). A person with such a conditional intention is under pressure not to consider acts incompatible with returning and is under pressure to adopt means of returning, where someone with an unconditional intention not to return is under pressure not to consider acts incompatible with *not* returning and is under pressure to adopt means of never returning. Rushlow, then, can be taken to be arguing, first, that his conditional intention not to return if his brother did not come home functioned differently from an unconditional intention never to return, and that his conditional intention to return if his brother came home exerted additional pressures different from those of the unconditional intender.

In *Rushlow*, then, the Mere Conditional Intent argument should be honored, since the relevant behavior was not guided at all by similar-deliberations but only by distinct-deliberations. This suggests another sort of case in which a verdict would be clear: cases in which the defendant's relevant behavior was guided exclusively by similar-deliberations. In cases of this sort, the defendant's conditional intention is sufficient for *mens rea*. Recall the case of *Irwin*, in which the Mere Conditional Intent argument was honored: Irwin took a hostage and threatened to kill her if he was not released from jail. Before it was clear, one way or the other, whether he would be released, he was disarmed. Since there is nothing to suggest that Irwin took any relevant action prior to taking his hostage—he did not, for instance, hide a weapon in his clothing, but instead took advantage of his guard's inadvertently giving him an opportunity to grab one—the only relevant stretch of conduct is that from the moment he took his hostage until the moment he was disarmed. During this period, he can have been

35. Thanks to Brandon Fitelson for pointing this out.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

guided only by reflection on scenarios that neither exclude nor include the condition governing the conditional intention. Plenty of the acts that Irwin performed were guided by deliberations in which his conditional intention places him under rational pressure to adopt means of killing his victim (that is, b3-deliberations). Included in the record, for instance, is the fact that in gripping the victim's neck and holding a knife to her throat, he applied so much pressure that the victim was lifted off the ground and lost her shoes. By making it that much less likely that his victim would struggle free, he adopted necessary means of killing her.

However, it is difficult to find any action that he performed that was guided by deliberations in which an unconditional intention would place him under pressure not to consider acts incompatible with killing his victim, but in which the conditional intention that he had placed him under no such pressure. An action guided by such deliberations would indicate that he was seriously considering courses of conduct that would make it impossible (relative to his beliefs) for him to kill his victim. But nothing that he did indicated this. Had he, for instance, thrown his victim aside and made a break for the door or done something else that indicated that he was considering such conduct, then he would have engaged in conduct guided by such deliberations. But as it happens, he did nothing like this; or if he did, he presented no evidence indicating that he did.[36] Thus Irwin's conditional intention played the same role in governing the deliberations that manifested themselves in conduct that an unconditional intention would have played, and so he satisfies the *mens rea* conditions for assault with intent to kill, contrary to the court's ruling.

The following general lessons can be drawn from these reflections. Here, an act counts as an instance of crime C only if S intends to A:

> S's conditional intention to A if X satisfies C's intent requirement *if* S's salient actions are guided exclusively by his similar-deliberations.
>
> S's conditional intention to A if X fails to satisfy C's intent requirement *if* S's salient actions are guided exclusively by his distinct-deliberations.

An action of S's that is included in the record is considered "salient" just in case it is not guided by reflection solely on scenarios that necessarily preclude the possibility of performing an act of type C. (The limitation to consideration of only salient actions is intended to exploit what is right

36. The defendant in United States v. Lake, 972 F. Supp. 328 (1997 U.S. Dist.) threatened to kill the victim if she did not give him the keys to her car, and from all appearances was ready to do so. But while the victim hesitated, he made an effort to grab the keys from her and eventually wrestled them from her grasp. While there is a question whether or not Lake actually had the conditional intention to kill the victim if she did not comply, it seems clear that if he did have that intention, he was still guided by deliberations in which he considered acts incompatible with killing his victim and he was hardly guided at all by deliberations in which he encountered pressure to adopt means to killing her.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

about the MPC test: Deliberative consideration of actions that are not salient is entirely innocent.) So we have identified a sufficient condition for a conditional intention's satisfying an offense's intent requirement and a sufficient condition for a conditional intention's failing to satisfy such a requirement. But a large class of cases fall into neither category. These are mixed cases in which part of what the agent does is guided by distinct-deliberations and part by similar-deliberations. *Holloway*, in fact, is a case in point. Holloway's conduct was never guided by reflection on X-scenarios—those in which his victim resists. His victim complied, and so he never acted on any conclusions that he reached about how to act in scenarios in which his victim did not comply; he may have thought about what to do in such scenarios, but his deliberations lay idle—they played no role in his conduct. So if there is a case to be made for the claim that Holloway's conditional intention is sufficient for the *mens rea* of carjacking, it must rest on the claim that before he came to believe that his victim would or would not comply, he was guided by deliberations in which he labored under pressures to adopt the means of killing her. In fact, for the period prior to and including the issuing of the threat, Holloway did indeed act on the results of such deliberations: During this period, he adopted all the necessary means for killing his victim (he procured a gun, he loaded it, perhaps he tried to envision killing the victim in order to make sure he would have the nerve, etc.). All of this conduct was guided by similar-deliberations.[37] But Holloway's salient conduct continued for some time after the moment at which the driver complied with the threat. In particular, it continued up until the moment at which he released her. This further behavior was guided entirely by deliberative reflection on not-X scenarios and thus by distinct-deliberations. Therefore Holloway differs from both Rushlow and Irwin in that part of his behavior is guided by deliberations that are structured by his conditional intention in just the way that an unconditional intention would structure them, and part of what he does is guided by deliberations in which his conditional intention plays an entirely different role from the role that would have been played by an unconditional intention. How should we decide whether or not Holloway satisfies the *mens rea* for carjacking?

Before proceeding to provide an answer to this question, note that at this point we have in place a framework for thinking about the relevance to *mens rea* of mental states different from but in some way similar to the unconditional intentions that are explicitly included in the law's definition of a particular crime. Intentions are of relevance to legal and moral responsibility because of the pressures that they impose on the deliberation guiding agents' behaviors. The relevance of other mental states, such as conditional

37. Notice that the action of issuing the threat was not dictated by Holloway's conditional intention to kill the victim if she did not comply but instead by his unconditional intention to steal the car. That action, then, is not relevant to the question of whether or not Holloway's conditional intention is sufficient for mens rea.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

intentions, derives from the similarities and differences between the pressures on the deliberation that guides the agent's salient conduct imposed by those mental states, on the one hand, and the pressures imposed by the particular intentions specifically invoked in the law, on the other. Theorists could agree on this much while disagreeing about where, exactly, to place the weight in mixed cases, like Holloway's, in which both similar- and distinct-deliberations guide some of the defendant's salient behavior. The account to follow, then, could be rejected together with acceptance of the framework developed here.

Still, anyone who accepts the general framework developed here for resolving the cases of relevance to us can agree that Holloway satisfies the offense's intent requirement just in case the deliberations structured similarly to those of the unconditional intender *play a more substantial role in guiding his conduct* than deliberations that are structured differently from those of the unconditional intender. Where theorists who follow this far may depart from one another is in their understanding of the notion of "substantial guidance." Or, to put the point more formally, we are in position to offer the following:

> *The Substantial-Similarity Test for the Mens Rea of Conditional Intent*: S's conditional intention to A if X satisfies C's intent requirement *if and only if* S's salient actions are guided more substantially by his similar-deliberations than by his distinct-deliberations.[38]

Here "salience" is defined as above so as to exclude consideration of actions guided by deliberative reflection on scenarios that necessarily exclude performance of the crime in question.

Although this test remains no more than schematic in the absence of an account of "substantial guidance," it nonetheless captures the results reached above with respect to *Rushlow* and *Irwin*: If the defendant's actions are guided exclusively by one set of deliberations or the other, then they are

38. One might wonder why any weight is given to the role of distinct-deliberations in guiding the defendant's behavior, given that similar-deliberations played some role. After all, we might say, if anything that he did was in response to deliberative pressures just like those of the unconditional intender, then his conditional intention played the same role as an unconditional intention, and he should be taken to satisfy the relevant mens rea requirements. The problem with this reasoning is that under it, gradations of responsibility are lost that it is important to maintain. After all, almost everybody has some mental state that plays some role in common with highly objectionable unconditional intentions. A person with an unconditional intention to kill at a particular time is under rational pressure to keep himself alive until then. Anyone who intends any innocent action at all at that time is under similar pressures, but it does not follow that his intention functions sufficiently like an unconditional intention to kill to warrant taking him to satisfy mens rea for, say, attempted murder. In order to see why not, we need to weigh the role that the intention plays in motivating other behaviors and compare that role to that played by an unconditional intention to kill. In appealing to the role of both the one sort of deliberation and the other, the Substantial-Similarity test makes possible the range of assessments that it ought to allow.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

guided more substantially by that sort of deliberation than by the other. So, at least in some cases, the test is relatively easy to apply. But to determine how it is to be applied in cases like Holloway's, more needs to be said about the notion of "substantial guidance" employed by the test.

One way to construe the notion of "substantial guidance" is as follows: To say that one sort of deliberation played a more substantial role in guiding an agent's behavior than another is to say that it is more tightly linked to the principles that unify the agent's behavior over time, where the "principles that unify an agent's behavior over time" are those facts about the agent's psychology by virtue of which his diverse bits of behavior are all interconnected. Often, although not always, an agent's acts are interconnected by virtue of their joint contribution to the attaining of some end at which the agent aims; what unifies the acts of, for example, reading law books, studying for exams, applying for internships, and paying tuition is the agent's desire to be a lawyer. In general, acts are unified by mental states that persist through performance of the acts in question or that make reference in their content to all of the various acts in question. When "substantial guidance" is construed in this way, the question before the court in cases like *Holloway* is this: Did the defendant's similar- or distinct-deliberations play a greater role in unifying the agent's behavior over time?

We can gain traction on this question by reflecting on the various evidential roles that conduct guided by a certain sort of deliberation can play in establishing an agent's mental state. However, to do so is not to mistake the way we know about the facts for the facts themselves. Rather, in the attribution of mental states, the way we know about the facts tracks the nature of the facts themselves; such attributions track the underlying principle that unifies the agent's behavior over time and can, therefore, tell us what those facts must be. To elaborate, notice, first, that we do not hesitate to ascribe to Irwin a conditional intention to kill rather than an unconditional intention to do so, despite the fact that his behavior is guided by deliberations just like those of a person with an unconditional intention. Why is this? Something other than the nature of the deliberations expressed by his conduct must convince us that Irwin's intention is merely conditional. We bring in information about what we assume to be Irwin's ends in acting. We assume, that is, that his only goal is to be released; a conditional intention serves that goal better than an unconditional, and so we ascribe to him a conditional intention or believe him when we hear him report that to have been his mental state.

This does not excuse him from responsibility, for the role played by his conditional intention was no different from that of an unconditional intention; he was guided, that is, only by deliberations in which he labored under pressure to adopt means of killing his victim—the same pressures that would have been encountered by an unconditional intender. Notice, however, that the assumption that leads to our ascribing to Irwin the conditional intention could be overridden. Say Irwin had been released and had

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

gone on to kill his victim. His further conduct would have led us to ascribe to him an unconditional intention to kill all along. His later conduct would have trumped the assumption about his goals that leads us, given that he was apprehended when he was, to ascribe to him a conditional intention. Conduct has evidential primacy in the assessment of mental state.

Now consider Holloway's action of releasing his victim. This conduct colors our account of the mental state that guided Holloway both at the time of the release and *prior* to his victim's compliance with the threat to kill her. While we might agree that Holloway's using a loaded gun indicates by itself that he was responding to deliberative pressure to adopt means of killing his victim, this leaves it open as to whether or not he has a conditional or unconditional intention, since either might lead him to use a loaded gun. To answer the question of which mental state he has, we need to bring in more evidence. Where in the case of *Irwin* we drew on assumptions about the defendant's overall goals in order to reach the conclusion that he was acting on a conditional intention, in *Holloway* we can draw on the more solid evidential foundation of further conduct: Since he released her later, he must have had a merely conditional intention when he issued the threat.

So the releasing of the victim has double-explanatory significance: It is an instance of conduct guided by distinct-deliberations—deliberations, in this case, about scenarios in which the victim complies—*and* it indicates that the agent's earlier conduct was guided by a conditional rather than an unconditional intention. By contrast, Holloway's earlier conduct—the loading of the gun and the other actions that were means of killing the victim—do not color our interpretation of his later conduct. They are part of what tells us what Holloway was doing—by expressing deliberation involving reflection on scenarios in which it is unclear if the victim will comply, they provide evidence for the claim that he had either a conditional or unconditional intention to kill rather than, say, an intention to bluff—but they do not tell us what he was doing over the full course of conduct included in the record. It is on these grounds that we should say that Holloway was more substantially guided by deliberations different from those of the unconditional intender than by deliberations similar to those of such a person. In this case, and frequently, the "principle that unifies the agent's behavior over time" is just the complex of mental states that includes the mental state of conditionally intending (it also includes the mental state of unconditionally intending to steal the car). In answer to the question: "Why did Holloway both load his gun *and* release his victim?" we are right to answer: "Because he conditionally, but only conditionally, intended to kill her." Since the act of releasing his victim plays a more important role in supporting this explanation than the act of loading the gun, the deliberative reasoning behind that act, structured by his conditional intention, plays a more substantial role in guiding his conduct. The result is that the case was decided incorrectly: Holloway's conditional intention does not satisfy the intent requirements of carjacking.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

The argument just offered for the claim that Irwin's but not Holloway's conditional intention is sufficient for *mens rea* is likely to encounter resistance, so it is worth expressing it more formally. The argument proceeds like this:

1. If distinct-deliberations guide a conditionally intending agent's behavior more substantially than similar-deliberations, then the agent's conditional intention is not sufficient for *mens rea*.
2. If the conduct guided by distinct-deliberations provides evidence for the agent's mental state over temporal interval $[t_0, t_3]$ and the conduct guided by similar-deliberations provides evidence for the agent's mental state only over temporal interval $[t_1, t_2]$ (where t_j precedes $t_{(j+1)}$ for all j), then the agent is more substantially guided by distinct-deliberations than by similar-deliberations.
3. The conduct of Holloway's that was guided by distinct-deliberations (e.g., his releasing his victim) provides evidence for Holloway's mental state over an interval beginning prior to his issuing the threat and ending with his release of his victim.
4. The conduct of Holloway's that was guided by similar-deliberations (e.g., his loading the gun) provides evidence for Holloway's mental state over an interval beginning prior to his issuing the threat and ending prior to his release of his victim.
5. So Holloway is more substantially guided by distinct-deliberations than by similar-deliberations.

Therefore Holloway's conditional intention is not sufficient for *mens rea*.

Premise 2 is not intended as part of an analysis of the concept of substantial guidance. It is rather intended as a claim to the effect that a fact about how we know a person's intention tracks a fact about the role that deliberation structured by that intention plays in providing an explanatorily unifying account of that person's behavior.[39] Notice that we cannot offer an argument similar to this on Irwin's behalf, because nothing that Irwin did was guided by distinct-deliberations, and so there is no deliberation-guided behavior to counterbalance the behavior guided by deliberations in which Irwin labored under pressure to adopt means of killing his victim.

Note, then, that the difference between Holloway and Irwin might be (although it might not be) a difference only in the time of apprehension.

39. In premise 2 and throughout the discussion of the argument of which it is a part, I am assuming that in ascribing mental states to a person on the basis of his behavior, we minimize attributing changes of mind to the agent. That is, explanations that attribute consistent states of mind to the agent over time are to be preferred to explanations that attribute mental states to the agent at one time that are abandoned at another. Or, to connect the point to our example, it is better to ascribe to Holloway a conditional intention all along, given that he released his victim, than to attribute to him an unconditional intention to kill her that he later reconsidered. There can be evidence for change of mind, of course, but in the absence of such evidence, appeals to change of mind should be minimized in the explanation of behavior. This principle is probably simply a corollary of Davidson's principle of charity: Ascribe mental states to others so as to maximize their rationality.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

If Holloway had been apprehended right at the moment when he issued the threat and before it was clear whether or not the victim would comply, then he would have satisfied the *mens rea* for carjacking and thus would have been guilty of attempted carjacking. By contrast, had Irwin been released, he very well might have allowed his victim to escape without killing her. What concerns the courts, quite legitimately, however, is all of the behavior in which the defendant engages that contributes to the alleged offense, and not behavior in which the agent might have engaged had circumstances been different. Just as we require that extenuating circumstances be actual rather than merely hypothetical if they are to ameliorate a defendant's legal responsibility, we require that deliberations distinctive of conditional intention manifest themselves in conduct if they are to be said substantially to guide the agent's behavior. Irwin, then, suffers perhaps from a form of bad legal luck, so to speak: Had he been released, he might have engaged in behavior substantially guided by deliberations like those that guided Holloway after the victim complied. Since he also might not have, we should not be bothered by this result.

The argument just offered for the claim that Holloway's conduct is more substantially guided by deliberations that differ from those of the unconditional intender than by deliberations in which similarity is to be found is not intended to be applicable to every case in which an agent with a conditional intention is guided by both similar- and distinct-deliberations. Ultimately, the task the court must perform in such cases is one of assessing the relative unifying roles of the two contrasting sets of deliberation. It is far from clear that a decision procedure for the guidance of such an assessment could be provided, or ought to be. Under the construal on offer, the concept of substantial guidance functions much like the legal standard of reasonableness. Many factors are involved in assessing whether or not, for instance, it would have been reasonable for a particular defendant to expect that his actions would cause a particular harm. In determining this, we expect judges to look to past decisions for guidance and to weigh various factors that are of obvious relevance to the question. But we do not expect judges' decisions on such matters to be strictly dictated. Similarly, the range of factors that contribute to the unifying role played by a form of deliberation would have to be weighed to determine what sorts of deliberations most substantially guided an agent's conduct, but we should not expect the judge's decision to be strictly dictated by those factors.

To construe the notion of substantial guidance as it has been construed here is to hold that the law has a distinctive concern with the principles that unify temporally extended conduct. Our moral assessments of agents are sensitive to the way in which we tell the full story of the agent's conduct. That story can paint the agent with a range of moral colors, but we recognize that a better story, a more accurate accounting of the agent's moral qualities, provides unity to the widest possible range of behaviors that the agent performs over time. The law has a similar concern. Although it remains an

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

open question whether the law's concern derives from morality's concern with the best explanations of temporally extended conduct or derives from some autonomous interest of the state, in either case the assessment of *mens rea* necessarily requires sensitivity to the unifying role of the underlying principles expressed by the sorts of deliberations that are structured by the defendant's intentions.

CONCLUSION

We care about intent, and we care particularly about intent that plays some role in guiding behavior, because we care about the nature of the deliberations expressed in the behavior of those whom we morally and legally assess; those underlying deliberations tell us how the agent directs himself through exercise of the powers that are distinctive of rational agency. To criminalize behavior that involves intent to a greater degree than we criminalize the same behavior absent the intent is to give a special place in legal responsibility to deliberative guidance. However, the fact that intentions can be conditional and can be thereby both like and unlike corresponding unconditional intentions forces consideration of the significance of intention to deliberative guidance and thus to legal and moral responsibility. The Mere Conditional Intent argument crystallizes the question: Under what circumstances is a conditional intention sufficient for the kind of deliberative guidance that makes an offense more severe, as stipulated by our laws?

The law has not treated this question consistently across jurisdictions or sometimes even within them and has been without adequate principled considerations for guiding its decisions. Part of the problem has come from the almost wholesale acceptance of the Model Penal Code's approach for resolving these difficult cases, which, as I hope to have shown here, is deeply flawed. Under the Model Penal Code's test, the question of whether or not a particular agent's conditional intention is sufficient for *mens rea* reduces to the question of whether or not the included condition ameliorates the evil of the intended act. The test tries to answer our question solely by comparing the content of conditional and corresponding unconditional intentions. By contrast, under the Substantial-Similarity test and the framework for assessing the relevance of conditional intention to responsibility of which the test is the upshot, an agent's intentions are taken to be of relevance to legal responsibility because of the structure they impose on deliberations that guide an agent's behavior and because of the role that such deliberations play in unifying explanations of the agent's behavior over time. The question, then, of whether or not a particular agent's conditional intention suffices for *mens rea* reduces to the question of whether or not that conditional intention imposes structure on the most explanatorily important of his deliberations in the same way as the mental states that the law explicitly bans.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press

In cases like *Rushlow* and *Irwin*, in which the agent is guided only by deliberations that are either unlike those that are prohibited or only by those that are like those that are prohibited, the Substantial-Similarity test saves the judge from hard questions by showing their answers to be irrelevant to the decision at hand. However, in cases like *Holloway*, in which some of what the agent does is guided by unacceptably structured deliberations and some by acceptable deliberations (at least by the standards of the law at issue), a judge must answer a difficult question. In such cases, judges must determine whether or not, all things considered, the actions that are guided just as they would have been by an unconditional intention play a more or less important unifying role than those in which the conditional intention played a distinctive role. We should not expect the answer to questions of this nature to be simple or necessarily codifiable. But nonetheless, the Substantial-Similarity test tells us when we need to answer a hard question and specifies that question as narrowly, perhaps, as the phenomena to be systematized allow.

https://doi.org/10.1017/S135232520404025X Published online by Cambridge University Press