Against Personifying the Reasonable Person

Matt King
University of Alabama at Birmingham

(forthcoming in *Criminal Law & Philosophy*)

1. The Counterfactual Test

The law often asks fact finders to determine if a defendant was reasonable in a particular respect. We might ask whether a defendant claiming self-defense, who believed that the victim posed an imminent and serious threat to their life, was reasonable in believing that.[1] Or we might ask of a defendant claiming provocation whether their extreme emotional distress was reasonable.[2] Or we might ask of a particular defendant, being tried over a negligence standard, whether their riskiness was reasonable.[3]

But how to decide such questions? One way in which fact finders are supposed to be guided is via a counterfactual test that personifies the reasonable person. First, we are to imagine the reasonable person being in the defendant's circumstances. The reasonable person has the defendant's evidence and faces the same background conditions. Then we are to determine whether the reasonable person would have done as the defendant did. Would the reasonable person have formed the same beliefs, made the same judgments, experienced the same level of anger or fear, or performed the same action? In short, we ask, "What would the reasonable person be like?" Answering that question is meant to point us to an answer about the defendant's

[1] E.g., NY Penal Code 35.15; MPC 3.09(2), Model Penal Code and Commentaries, American Legal Institute.
[2] E.g., MPC 210.3(1)(b) or *Girouard v. State*.
[3] MPC 2.02(2)(d)

reasonability in the circumstances. If the defendant was like the reasonable person, then the she was reasonable.[4]

For example, suppose Chester has clearly killed the victim, but he acted under circumstances that he believed "to provide a moral justification or extenuation for his conduct".[5] Perhaps Chester believed that the victim had been following him home from the metro stop, intent on killing him. Chester thought that the victim was a personal enemy of his based on the similarity of dress and gait of the victim, though he never got a good look at his face. Finally, just before Chester shot and killed the victim, he saw a glint of metal in the victim's hand, which he took to be a weapon of some kind. It simply never occurred to him that it might just be the victim's keys (which it was).

Chester may have sincerely thought that lethal force was necessary to defend himself. But it will be insufficient for Chester to show that he genuinely believed he had such a justification. Additionally, he must be *reasonable* in that belief.[6]

To determine whether he was so reasonable, the jury might well be tasked with imagining whether the reasonable person would have believed as Chester did given the same circumstances.[7] Would the reasonable person have concluded, based on the same visual evidence and background information, that the victim posed an immediate threat of harm and that resisting

---

[4] Cf. Peter Westen, "Individualizing the Reasonable Person in Criminal Law" *Criminal Law and Philosophy* 2: 136-162 (2008): "[I]nstead of asking, 'Was the actor's use of force reasonable', *one can just as well ask*, 'Would the reasonable person have used the force the actor employed?' A reasonable person is reasonableness rendered incarnate" (139, my italics).

[5] MPC 210.6(4)(d).

[6] Importantly, this is not all that would have to be true. For instance, even if Chester's belief were reasonable, his response would have to be proportional to that threat. Here I am only concerned with the reasonability component of the defense of self-defense, however, not the defense itself.

[7] Here is just one example: "The statute requires, however, that the defendant's belief be reasonable, and not irrational or unreasonable under the circumstances; that is, *would a reasonable person in the defendant's circumstances have reached that belief*," (State of Connecticut Criminal Jury Instructions, 2.8-1: Self-Defense and Defense of Others -- § 53a-19, my italics).

with force was necessary in order to protect oneself? Chester's beliefs are reasonable so long as the reasonable person would have believed the same thing. Otherwise, he fails to reasonably believe, and his claim of self-defense ought to be rejected.

Despite its prevalence in jurist opinion and jury instruction, this counterfactual test is a hopeless guide to determining defendant reasonability. Indeed, it *cannot* give guidance on the question it is meant to address. Moreover, once we attend to these difficulties with the counterfactual test, we will see that there is reason to abandon the personification of the reasonable person as well.

2. What the Reasonable Person Is Not

Why does the test fail? To begin, note that it's either the case that the reasonable person can be unreasonable in some respect or it isn't. If it is possible, then we could never use the counterfactual test to tell us whether the defendant was reasonable in that respect. For if the reasonable person could be unreasonable in some respect, then for all we could discern the defendant could in fact be exactly like the reasonable person. In such cases, the test wouldn't even be of the right *sort* to give us the answer we are after. But the point of the test is to give us guidance on whether the defendant was in fact reasonable in some respect. This rules out the possibility that the reasonable person could be unreasonable.

To some too quick this might seem too quick.[8] We might think that the reasonable person is nonetheless capable of being unreasonable in certain ways. For example, we might interpret the role of reasonableness in the provocation defense in this way. One who has access to the

[8] My thanks to an anonymous referee for pushing me on this score.

defense of provocation has been reasonably angered or has reasonably lost control of themselves temporarily, but they don't *act* reasonably should they kill their provoker as a result.[9] If it were reasonable to follow through on one's reasonable provocation with killing, then we should expect provocation to serve as a more perfect justification. But this is not how the defense of provocation works. It doesn't render one's action reasonable; it merely reduces one's crime to manslaughter. The reasonable person, however, is never guilty of manslaughter. It may in some sense be more understandable to kill out of provocation than out of a desire to gain one's inheritance or out of pure envy, but it is not something the reasonable person would do. This suggests that, as far as provocation goes, we are only interested in the reasonable person to the extent that they are meant to guide us with respect to the loss of control or extreme emotional disturbance.[10] So we may ask of the defendant: were they reasonable in losing control in the face of legally adequate provocation? We might then consider whether the reasonable person would have similarly lost control. But we need not ask whether the reasonable person would have then killed the defendant as a result. It is not reasonable to kill even when provoked.[11]

Thus, whenever we inquire about the reasonable person, we are always interested in their reasonability *in some particular respect*.[12] We might inquire as to the reasonable person's

---

[9] Cf. John Gardner, "The Mysterious Case of the Reasonable Person," *The University of Toronto Law Journal*, 51(3): 273-308 (2001): "To benefit from the provocation defence, the defendant must have 'reacted reasonably' only in the sense that he must have been justified in losing his temper to the point at which he was apt to kill. Naturally this does not mean that he was justified in killing" (273).

[10] The MPC opts for a more inclusive category to the legitimate grounds for provocation in traditional common law (MPC 210.3(1)(b)).

[11] That this is a settled question is supported by the observation that fact finders are not asked whether the reasonable person in the defendant's circumstances, having been provoked, would have reacted by killing the defendant. Other unreasonable action is possible from reasonable beliefs or emotions besides provoked killings. That the defendant was reasonable in believing their safety to be in imminent danger doesn't guarantee that they were reasonable in using deadly force on their attacker. The defendant's use of force may have been disproportionate or they may have had a legal duty to retreat.

[12] Some interpret 'reasonable' in the law as synonymous with 'justified' (John Gardner, "The Many Faces of the Reasonable Person," *Law Quarterly Review* 131 (2015)). On this interpretation, it is easy to see why we would be interested in the 'justified person' only to the extent that we are inquiring about their justification. Was the

beliefs, or attitudes, or fear, or, indeed, emotional disturbance. It is impossible for the reasonable person to be unreasonable in the respect relevant to the inquiry. This is compatible with a particular defendant being reasonable in that respect and yet whose conduct is nonetheless unreasonable (as provocation illustrates). But the reasonable person is never relevantly unreasonable. To entertain that possibility is merely to shift our attention from one framework, for example, fear, to another framework, action. In either respect, we may ask about a defendant's reasonableness, and so consider whether her fear or her action was reasonable. In each case, we might be asked to consult the reasonable person, but each consultation would be limited to the relevant respect of inquiry: fear or action.[13] For the reasonable person neither fears unreasonably nor acts unreasonably, though it is certainly possible for actual people to fail to maintain their reasonableness from fear to action.[14]

Thus, the reasonable person is never unreasonable. This may seem a trivial exclusion. But it is important to make note of it, at least because it shows why one interpretation of the personification of the reasonable person is ruled out. In some commentaries and models, the personification of the reasonable person is made using the ordinary person or the "person of

---

defendant's belief justified? Was their fear? Was their action? All are separate respects in which we might look to the balance of reasons. Interestingly, if one takes reasonability to mean no more than justification, it turns out the counterfactual test is even more clearly problematic (see n. 18)

[13] Thus, I think it a mistake to suppose that the reasonable person, on any interpretation of reasonability, could get things wrong in the relevant respect (cf. Gardner, "The Many Faces of the Reasonable Person"). Though the reasonable person and a defendant could be perfectly identical in their fear or belief of imminent attack, we should not therefore conclude that the reasonable person could be capable of going on to do something unreasonable as a result. To think the reasonable person could do that is simply to supplant them with a different imaginary placeholder – perhaps *a* reasonable person – once the relevant question has been settled (e.g., whether the defendant's belief was reasonable).

[14] Additionally, note that if the reasonable person could be such so as to act unreasonably from nonetheless reasonable fear, then being like the reasonable person would not necessarily be to the defendant's credit, contrary to its usual purposes. Granted, it may better to have acted unreasonably on reasonable fear rather than unreasonable fear, but in that event the only relevant consideration is the reasonableness of the fear, which, I maintain, is the only fear of which the reasonable person is capable.

ordinary prudence" as a guide.[15] Read in one way, the ordinary person is just a statistical metric. We are to imagine a person who is like most citizens: not especially careful, observant, or intelligent, but also not especially risky, oblivious, or dim-witted. The counterfactual test then asks us to take the ordinary person and imagine how they would have acted under the circumstances.

Critical commentators have already noted that such a standard would get something importantly wrong about the role of reasonableness in the law.[16] Being unreasonable does not necessarily require diverging from the norm, understood as a statistical measure. It could be that most folks in the defendant's circumstances would have come to believe or act as he did. All that would show is that most people could have believed or acted unreasonably.

This observation is confirmed by excluding the possibility of the personified reasonable person being unreasonable in any respect. For, patently, defendants are most often ordinary people. It is no precondition on finding oneself on the wrong side of the law that one is especially vicious in any particular way. If our defendant is an ordinary person, however, we could not personify the ordinary person and come to a judgment that such a person would not be like the defendant. For the defendant is likely to be just such an ordinary person. And if the defendant could be an ordinary person, and the defendant believed and acted as he did, then we have straightforward evidence that ordinary people can be like the defendant. Indeed, that they can be exactly like the defendant in the relevant respects.

Similarly, if we allow that reasonable persons can be in some respect unreasonable, then we could not conclude that the defendant was unreasonable by personifying the reasonable

[15] The latter is especially common in tort law negligence standards to determine whether the defendant took due care, understood as reasonable care.

[16] Cf. Westen, "Individualizing the Reasonable Person," 138.

person in her circumstances. We would gain nothing, since the reasonable person might have been unreasonable in precisely the way in which the defendant was.

Thus, if it were possible for the reasonable person to nonetheless be unreasonable, then a finding that the reasonable person would have done as the defendant did would not settle the proper question for us. A positive finding would not imply that the defendant was in fact reasonable under the circumstances. In short, if the reasonable person could themselves be unreasonable in some fashion, the counterfactual test is not able to do its guiding work.

3. Guidance and Grounding

So, the only way for the test to do the requisite work is by excluding the possibility that the reasonable person can be unreasonable in any respect. We might say that, by definition, the reasonable person is never unreasonable. It just is what it is to be the reasonable person that one is always reasonable.

But on such a requirement the counterfactual test is still problematic. The problem for it mirrors a classic objection to a particular (perhaps naïve) formulation of virtue ethics. Suppose we want to know how to act virtuously in a given situation. What is the virtuous thing to do? Well, we might imagine the virtuous person, the one disposed to do virtuous things, and ask what would they do in the circumstances.[17] In order to know what the virtuous person would do, however, we'd have to know what makes them virtuous. But if we knew that, we'd have already

[17] Its important that the guide is *the* virtuous person and not *a* virtuous person. The former is an ideal, whereas the latter is a particular embodiment. Only the former is meant to be instructive. As will become clear, the latter is also of no help, since in order to identify a particular agent as a virtuous person we would first have know something about the ideal.

determined what virtue requires, which was the very question the counterfactual test was meant to help us with.

Matters are the same in thinking about the reasonable person. What would the reasonable person believe or do under these circumstances? In order to answer such a question, it seems we'd need to know what makes the reasonable person reasonable. But knowing that would eliminate the need for appealing to the reasonable person in the first place.

Consider Chester again. Chester believed that there was a personal enemy of his following him, who wished him serious and immediate harm, and who was drawing a weapon from their pocket. He formed most of these beliefs on the basis of fairly sketchy observational evidence. Using the counterfactual test, we might ask whether the reasonable person (RP), on the basis of seeing a person behind them at several points in one's walk home, and without ever seeing the person's face, would come to the conclusion that it was a particular individual and that he was following RP home? In answering that question, we are guided only by the knowledge that the reasonable person is never unreasonable. So if it would be unreasonable to draw that conclusion, then RP would not have done so. If it would be reasonable to do so, however, then RP might have.[18] We might wonder whether it is ever reasonable to identify persons without a clear view of their face, but, of course, there can be other significant features that might be sufficient grounds for drawing a suitable inference. We might suppose that there is some epistemic standard for inferring certain facts on the basis of certain visual experience under the right conditions, such that we can fit the circumstances in Chester's case to that standard. As jurors, we might put ourselves in the circumstances and consider what beliefs we would have

[18] Recall those that take 'reasonable' to mean no more than 'justified'. How could we determine whether the defendant was justified (in some respect) by considering what the justified person would have done (or believed or feared, etc.)? That question cannot be usefully interpreted in any other way than, "Was the defendant justified (in the relevant respect)?" But since that is the very question the counterfactual test is meant to give guidance on, it cannot give meaningful guidance as to whether an individual was in fact justified.

formed. But we are likely just ordinary people, so able to be unreasonable as much as Chester himself.

The point here is that, whatever our imaginative capacities, we have little to go on in trying to assess the counterfactual question regarding the reasonable person. To determine whether RP would conclude, on the basis of a glint of metal, that the victim was pulling a weapon from his pocket, it helps not at all to imagine RP in the circumstances, without having some grip already about what reasonableness requires under those circumstances. We could confidently claim that the reasonable person would not have believed as Chester did only if we already know what is reasonable to believe in the case.

But that means that in order to determine what the reasonable person *would* have done in the defendant's circumstances, we'd already know enough to determine whether the defendant was theirself reasonable, without ever concluding the counterfactual test.

Relatedly, there's another way in which the counterfactual test fails to give guidance. Suppose we *disagree* over whether the reasonable person would have acted as the defendant did. I say that they would because the reasonable person would have some quality X, whereas you say they would not because the reasonable person has some quality Y. And suppose further that we agree that X and Y are incompatible. Now, to settle which characterization of the reasonable person is correct, we'd have to look at whether having quality X or Y (or acting from those qualities, or whatever) would be reasonable. But that is the very question for which we looked to the counterfactual test for guidance in the first place. If the test can't adjudicate between competing claims, it again fails to guide.

Without knowing something about reasonability itself, we simply have no guidance on what the reasonable person would do. But since the counterfactual test is supposed to be a way of

getting at what reasonability involves, it only produces answers when it presumes the very thing it's after. Such a procedure is hopeless.

4. Implications and Complications

How does a hopeless procedure nonetheless get results? After all, juries are frequently asked to determine a defendant's reasonableness and are given the counterfactual test for guidance. Such juries are not stumped by the question. On the contrary, they come to answers. One might take the ordinariness of this routine as evidence that the test is useful.

But to show that the question is asked and that juries render verdicts does not show that the juries answered the question. Of course, juries render judgments about the reasonableness of the defendant. But there are a variety of possibilities regarding this fact. First, they may try to judge independently what reasonableness requires. In this case, the counterfactual test serves no actual purpose, and they (either explicitly or implicitly) ignore that element of the instruction.

Second, they may be substituting themselves or an "ordinary" person into the counterfactual test. They simply try to determine whether they could see themselves (or someone they know) coming to the defendant's beliefs, or acting as the defendant did, under those circumstances. While the law may reject the statistical interpretation of the reasonable person, jurors may turn to less ideal imaginings all the same.

Third, if they do take the instruction seriously, they may try to imagine how one would behave, but fail to employ the idealized reasonable person. Not knowing what reasonability requires, they would look for a surrogate, perhaps someone they think reasonable (i.e., they use "a" reasonable person).

Finally, it could of course be the case that they never really move from judgments about reasonability to judgments of culpability, but the other way around.[19] If they think the defendant acted wrongly and is probably guilty overall, they may be moved to remain consistent and claim that the defendant was unreasonable.

What juries in fact do is an empirical question. It could of course be the case that they do manage to first answer the question of whether the defendant was reasonable or not. My claim here is not that they cannot answer such a question, merely that they cannot do so *by employing the counterfactual test*. Instead, a plausible conjecture is that, if they answer the question, they answer it more or less directly. Actually, the conjecture helps elaborate on why the counterfactual test is hopeless. What the reasonable person would be like in particular circumstances follows from what is reasonable in those circumstances, rather than the other way around. So, juries can answer both questions. But since the answer to the counterfactual test depends on the answer to the question of what is reasonable, the counterfactual test is logically dependent on a prior question. And that question is as answerable, indeed, I think more easily answered, than is the counterfactual test.

Importantly, however, it means we offer no guidance to fact finders by directing them to the counterfactual test. If we want to offer them additional guidance, beyond their own grasp of reasonability, we should look elsewhere.

One might conclude that, since fact finders come to the relevant conclusions, even if the counterfactual test offers no guidance, there is no harm in its appeal. But careful reflection on the problems with the test itself casts a shadow on personifying the reasonable person at all.

[19] Cf. The so-called "Knobe effect" (Joshua Knobe, "Intentional Action and Side Effects in Ordinary Language," *Analysis* 63: 190-194 (2003).)

To see why, note that reasonableness applies to a spectrum of possibilities. There are a variety of things one may reasonably believe or do under particular circumstances. If that's right, then it will be possible for the reasonable person, however conceived, to have believed and done differently than the defendant, without ensuring that we can infer that the defendant was thus unreasonable.[20]

These observations, however, reveal a striking conclusion. They suggest that our primary concern is not whether the defendant was reasonable, but whether they were *un*reasonable. Thus, our interest should not be in what the reasonable person would do, but in what the reasonable person could do. *Could* the reasonable person have believed as Chester did? The reasonable person may still have believed differently, since her reasonability only guarantees her beliefs fall within the relevant spectrum. But if she couldn't believe as Chester did, then we can conclude that Chester believed unreasonably.

Once we've fixed the question in terms of a capacity to believe while remaining reasonable, however, it is apparent that the question we're really asking is simply whether those beliefs were reasonable. The counterfactual test has evaporated. Indeed, the surrogate question is no longer a counterfactual at all. It is not an assessment of what would transpire under the circumstances if the reasonable person were involved, but what the reasonable person is capable of. Given that the only thing the reasonable person is necessarily incapable of is being unreasonable, it is plain that that test is not a guide to the question of reasonability at all.

---

[20] Compare the Homicide Act of 1957, from English criminal law:

> "Where on a charge of murder there is evidence on which the jury can find that the person charged was provoked…the question whether the provocation was enough to make a reasonable man do as he did shall be left to be determined by the jury; and in determining that question the jury shall take into account everything both done and said according to *the effect which…it would have on a reasonable man*" (my italics).

The question posed presumes there is *an* effect on the reasonable person some such action or words would have. But this is far too narrow a supposition. Surely, the reasonable person *could* resist being provoked by some reasonably provocative trigger, even if it is not unreasonable to fail to so resist.

Moreover, once the question concerns the defendant's reasonability directly, there is no further purpose to which to put the reasonable person personified.[21] Irrespective of any difficulties such a personification may involve, the argument here concludes that we can gain no useful guidance from consulting it.[22]

[21] At least, no theoretical purpose. Phrasing a jury instruction or rule of law in a particular way could be shown to be instrumentally valuable. For example, by doing a better job of getting fact finders to attend to the appropriate considerations or avoid systematic error. But even if this were the case, the instruction or rule would not be *guiding* the fact finders, and similar advantages could equally attach to all manner of rationally unconnected processes or methods.

[22] My thanks to Gideon Yaffe, Scott Shapiro, and the dinner group discussants at Yale University for their helpful feedback.