

How a pure risk of harm can itself be a harm: A reply to Rowe

H. Orri Stefánsson

The question of whether a pure risk of harm can itself be a harm is currently under debate.¹ By a risk of harm, I shall simply mean a positive probability of a harmful event.² A *pure* risk of harm, however, is a risk of harm that does not actually lead to the harmful event in question, nor are the potential victims of the harm aware of the risk (Oberdiek 2017: 67; Rowe 2021: 695). To take an example, suppose that unbeknownst to Beth, she has a heart condition that could be fatal under sufficient stress. Luckily, she never experiences such stress, and instead lives her whole life without this condition affecting her in any way and without her ever finding out that she has the condition. Then Beth suffers (merely) pure risk of harm, due to this heart condition of hers.

How should we then understand *harm*? Although far from being universally adopted, I shall in what follows assume Feinberg's (1987) account according to which harm is a setback to a legitimate interest. Why that account? Because it has recently been adopted by all parties to the debate about whether pure risk of harm can itself be a harm.³

Rowe (2021) raises important objections to the claim that pure risk of harm can itself be a harm. He calls this claim the *Risk Harm Thesis*. In line with the literature on the ethics of risk imposition, he considers two accounts of 'risk'. According to the *subjective* account, the probabilities that figures in claims such as 'Beth is at a risk of suffering a heart attack' are the degrees of belief of some person, usually someone who is appropriately influenced by the evidence we would expect a 'reasonable person' to possess (see, e.g., Oberdiek 2017: ch. 2). According to the *objective* account, however, the probabilities in question are mind independent. As Rowe points out, objective probabilities are typically understood as *relative frequencies* in the literature on the ethics of risk (e.g., by Finkelstein 2003 and Oberdiek 2017). Thus understood, the claim that Beth faces a 0.1 risk of suffering a heart attack means that 10% of people that are relevantly similar to her—10% of her reference class, as it is often put—suffer a heart attack.

To illustrate the problem with the claim that pure *subjective* risk of harm could itself be a harm (a claim defended by Oberdiek 2017), Rowe uses the following example:

¹ For some canonical examples, see Perry (1995), Finkelstein (2003), Oberdiek (2017), and Rowe (2021).

² The risk can vary in magnitude in proportion to both the probability and the magnitude of the possible harm.

³ See in particular Finkelstein (2003), Oberdiek (2017), and Rowe (2021).

(Russian Roulette Aggressor) Completely unbeknown to Adam, who is enjoying the weather on a park bench, Beth is aiming a six-shooter at him with the intention to kill him. Beth loads a single bullet into the cylinder. However, Beth in fact loads a blank, but does not realize this. Reasonably believing that there is one bullet and five empty chambers in the cylinder, Beth pulls the trigger. (Rowe 2021: 696)

What this example shows, Rowe thinks, is that the claim that pure subjective risk could be a harm is vulnerable to what he calls the *interference objection*. ‘A subjectively imposed risk is unable to set back an agent’s interests, because a mere subjective risk of harm cannot itself interfere with the interests of the victim in order to set these interests back’, Rowe (2021: 697) points out. ‘This is because the subjective risk exists only in the mind of Beth. A credence itself cannot plausibly set back anything. Therefore a subjective risk of harm is itself not a harm.’

To illustrate the problem with the claim that pure *objective* risk of harm could itself be a harm (a claim defended by Finkelstein 2003), Rowe uses the following example:

(Deadly Bingo) Beth has a sophisticated lottery ball machine that triggers a loaded gun the moment ball number 13 is drawn. The gun is pointing at Adam, who is sitting on a park bench blissfully unaware of Beth’s contraption. Once the machine is turned on, twenty lottery balls spin in the container. (Rowe 2021: 698)

To build his case against pure objective risk being harmful, Rowe asks *when* Beth begins to harm Adam. (Below I refer to this as the *timing objection*.) As he points out, it seems unreasonable that Beth would start to harm Adam before she turns on the machine. Moreover, the proponent of the objective Risk Harm Thesis can hardly say that Beth starts harming Adam once ball number 13 is drawn. For then Adam will in fact be harmed, and we are not dealing with *pure* risk. Instead, it would seem that we must say that if Beth harms Adam in virtue of pure risk imposition, then she starts harming him the moment she pushes the ‘start’ button.

The problem with saying that Beth starts harming Adam the moment she pushes the button, Rowe thinks, can be illustrated by imagining that ‘a perfect predictor, knowing all the antecedent causes, could predict whether ball number 13 would be drawn when Beth turns on the machine’ (Rowe 2021: 698). Suppose Beth pushes the button and the predictor predicts with certainty⁴ that ball number 13 will not be drawn. ‘Does the objective risk nevertheless harm Adam?’

⁴ Given that the predictor is perfect, it may be unnecessary to qualify that he predicts this *with certainty*. Still, I add this qualification since Rowe’s lesson from the example would clearly not hold if the predictor were not sure about his prediction.

Rowe (ibid.) asks. That seems implausible, he thinks. ‘A rational guardian of Adam’s interests who is aware of the perfect prediction (while Adam himself is not) would not be willing to pay anything to remove the risk itself, because the risk will not result in Adam being shot. But if an objective risk of harm is itself a harm, then, other things being equal, it would be rational to want to remove it.’ (Rowe 2021: 699)

The argument in the last paragraph may not make much sense for a risk that arises from an indeterministic process. In a deterministic world, however, Rowe’s objections against the Risk Harm Thesis work well when the risk in question is either understood subjectively or in terms of relative frequencies (which, as the reader will recall, is the account of objective risk that Rowe assumes). In contrast, I shall argue that his objections can be met if we understand the probabilities involved in risk claims as the degrees of belief of an *ideal* reasoner who knows *all* the relevant facts.⁵ To make my argument as charitable to Rowe as is possible, I shall not make any assumptions that are inconsistent with determinism. Nevertheless, I shall suggest that the account of probabilities I am proposing makes it plausible that pure risk of harm could itself be a harm whether we live in a deterministic or indeterministic world.

Let me first briefly explain the account of probabilities on which my argument will be based. Bradley has recently argued that objective probabilities (or chances) should in general be understood in terms of the degrees of belief of an ideal reasoner who is ‘as well-informed about all the relevant facts as it is possible to be’ (Bradley, 2022: 2). In other words, objective probabilities express the ‘best-possible judgement that can be made, given these facts’ (ibid.).

Now suppose that Beth faces a risk of heart attack where the relevant probability is objective in the above sense, that is, they are the degrees of belief of an ideal reasoner who is aware of all the relevant facts. This risk would presumably not be subjective in Rowe’s sense, since it would not exist *only* in someone’s mind (Rowe 2021: 697). In fact, it might not exist in *anyone’s* mind. Still, we can ask whether the Risk Harm Thesis, with these objective probabilities, would be undermined by Rowe’s interference objection. Can a risk thus understood interfere with someone’s interest? I think it can.

Recall that the reason why it seems implausible that pure subjective risk could interfere with Adam’s interest in the Russian Roulette Aggressor, is that the risk in question exists solely in Beth’s mind. In general, one might think that something that exists only in Beth’s mind could not interfere with Adam’s legitimate interests.⁶ But then since objective risk, understood the way I am

⁵ Rowe’s objections could be similarly met, I contend, with the help of ‘Humean chances’ as developed by for instance Frigg and Hoefer (see, e.g., their 2010 and 2015; see also Hoefer 2007).

⁶ There do however seem to be exceptions. Perhaps Adam has a legitimate interest in Beth loving him even though Beth’s love would only exist in her mind. But this does not raise a problem for my argument. After all, I am not

proposing, does not only exist in someone's mind, it does not face the interference objection. In fact, objective risk, thus understood, is not much different from, say, temperature. Mind-independent facts completely determine both quantities. And just as the fact that the temperature has reached a certain degree might interfere with Beth's legitimate interests—it might for instance make it unsafe for her to go for a run—so the fact that she faces a particular risk of heart attack might interfere with her legitimate interests, for instance, her interest in safety (even if she does not know about the risk).

What about Rowe's timing objection? I think that objection can be met too, given the above understanding of objective risk. In *Deadly Bingo*, I contend that Beth harms Adam as soon as she pushes the 'start' button, assuming that doing so makes it objectively possible—that is, creates an objective probability greater than zero—that ball number 13 will be drawn and that Adam will be shot. What should we say then about the case when, after Beth pushes the button, a 'perfect predictor' (Rowe 2021: 698) predicts with certainty that ball number 13 will in fact not be drawn? If a perfect predictor really can predict this with certainty, then the objective probability of ball number 13 being drawn is zero, given how I am suggesting we understand objective probabilities. So, Adam faces no (objective) risk. But if Adam faces no objective risk in this example, then the example cannot establish that (pure) objective *risk* of harm cannot be harmful.

To see why Adam faces no objective risk, given how I am suggesting we understand the relevant probabilities, recall that I am proposing that we understand objective probabilities as the best-*possible* judgement that can be made. If the predictor in question is indeed 'perfect', as Rowe assumes, then perfection is possible. Therefore, the best-possible judgement will be perfect. In other words, the objective probabilities (i.e., the best possible judgements) will correspond to the predictor's judgements. So, if he predicts with certainty that ball number 13 will in fact not be drawn, then the objective probability of ball number 13 being drawn is zero. Thus, Adam faces no (objective) risk. On the other hand, if—contrary to Rowe's and my assumption—the predictor is not perfect, then Adam may well be facing objective risk, and we have not yet seen a reason why a 'rational guardian of Adam's interests' (Rowe 2021: 699) would *not* be willing to pay something to have the risk removed (even though Adam is unaware of the risk).

To conclude, I hope to have shown that the objective version of the Risk Harm Thesis can be defended against Rowe's objections, if objective risk is understood in terms of the degrees of belief of an ideal reasoner who is as informed of all the relevant facts as it is possible to be. Further, I think the thesis is independently plausible, when objective risk is thus understood. Saying

claiming that something that exists only in Beth's mind couldn't interfere with Adam's legitimate interests. Instead, I will argue that even if Rowe is right in that pure subjective risk cannot itself be a harm, pure objective risk—understood as I am suggesting—can itself be a harm.

that someone faces objective risk can then strictly speaking be reduced to a claim about certain mind-independent facts obtaining. So, it should not be particularly mysterious that someone could be harmed by pure objective risk. Finally, note that this holds irrespective of whether the world is deterministic or indeterministic. Pure risk of harm can thus itself be harmful, whether or not we live in a deterministic world.⁷

Stockholm University,
Swedish Collegium for Advanced Study,
and Institute for Futures Studies (Sweden)
orri.stefansson@philosophy.su.se

Funding Financial support from Riksbankens Jubileumsfond [Pro Futura Scientia XIII] is gratefully acknowledged.

References

- Bradley, R. 2022. Conditional chance and warranted credence. *London School of Economics Working Paper* (available at <https://personal.lse.ac.uk/bradleyr/>).
- Finkelstein, C. 2003. Is risk a harm? *University of Pennsylvania Law Review* 151: 963–1001.
- Feinberg, J. 1987. *The Moral Limits of the Criminal Law Volume 1: Harm to Others*. New York: Oxford University Press.
- Frigg, R. and Hoefer, C. 2010. Determinism and chance from a Humean perspective. In *The Present Situation in the Philosophy of Science*, eds. D. Dieks, et al., 351–271. Springer.
- Frigg, R. and Hoefer, C. 2015. The best Humean system for statistical mechanics. *Erkenntnis* 80: 551–574.
- Hoefer, C. 2007. The third way on objective probability: A sceptic’s guide to objective chance. *Mind*, 116: 449–496.
- Oberdiek, J. 2017. *Imposing Risk: A Normative Framework*. Oxford: Oxford University Press.
- Perry, S. 1995. Risk, harm, and responsibility. In *Philosophical Foundations of Tort Law*, ed David G. Owen, 321–6. Oxford: Oxford University Press.
- Rowe, T. 2021. Can a risk of harm itself be a harm? *Analysis* 81: 694–701.

⁷ I am grateful to Richard Bradley for illuminating correspondence about the topic of this paper, and to two referees for this journal for very helpful comments and suggestions.