

Commentary to: Bermúdez, J.J. (2022). Rational Framing Effects: A Multidisciplinary Case. Behavioral and Brain Sciences. DOI: <https://doi.org/10.1017/S0140525X2200005X>

Rational framing effects and morally valid reasons

Tomasz Żuradzki

Jagiellonian University in Kraków (Poland), Institute of Philosophy & Interdisciplinary Centre for Ethics, ul. Grodzka 52, 31-007 Kraków, Poland, t.zuradzki@uj.edu.pl

Abstract: I argue that the scope of *rational framing effects* may be broader than Bermúdez assumes. Even in many "canonical experiments," the explanation of the judgment reversals or shifts may refer to reasons, including moral ones. Referring to the Asian Disease Paradigm (ADP), I describe how non-consequentialist reasons related to fairness and the distinction between doing and allowing may help explain and justify the typical pattern of choices in the cases like ADP.

José Luis Bermúdez (2022) contrasts simple cases of framing effects (e.g., the Asian Disease Paradigm – ADP) where "frames prime responses" with more complicated situations where "frames can function reflectively, by making salient particular reason-giving aspects of a thing, outcome, or action." He concludes that the focus on these simple cases "has blinded us to the existence of *rational framing effects*." I agree with Bermúdez that, surprisingly, most research in moral psychology, behavioral economics, and experimental ethics on framing effect concentrate on the very existence and the scale of this effect while neglecting its explanatory and justificatory dimension. In particular, many studies neglect to investigate valid reasons people may have to reverse or change their judgments regarding morally salient choices.

For example, Tversky and Kahneman (1981) seemed to be fully aware of the ethical implications of their findings ("When framing influences the experience of consequences, the adoption of a decision frame is an ethically significant act"). Therefore, it is surprising that most research on ADP does not even mention an ethical dimension of this choice situation, although it consists of making a life-saving decision. This observation also applies to Bermúdez's works (2021, 2022), which seem to treat the standard pattern of choices in "the canonical experiments" (incl. ADP) as irrational. Although

Bermúdez discusses some highly stylized moral reasons in the examples of Agamemnon and Macbeth, he mentions more morally relevant examples (i.e., abortion) in merely one sentence.

I assume that two types of moral non-consequentialist reasons may be particularly helpful in *explaining* and *justifying* the standard pattern of choices in the cases like ADP: fairness (egalitarian reasons) and the distinction between doing and allowing. First, programs B and D ("1/3 probability that 600 people will be saved [nobody will die], and 2/3 probability that no people will be saved [600 people will die]) are fully egalitarian (cf. Segall 2016). Not only has every individual ex ante equal chance for survival (1/3), but there is also full equality ex post: everyone is alive or dead. In turn, programs A and C (200 people will be saved [400 people will die]) are only partially egalitarian. Even if one understands these two options as cases of random distribution (so again, every individual has ex ante equal chance for survival, i.e., 1/3), there is inevitable ex post inequality (some will be alive and some dead).

Second, the next non-consequentialist dimension of this choice concerns the distinction between doing and allowing (Woollard 2015). Programs A and B in the ADP are formulated in terms of "people will be saved," so the reference point is 600 people dead (or precisely speaking, "almost dead"). Therefore, many may assume that the choice concerns the distribution of benefits since the harm has already been done, and choosing program A is actively saving ("doing something for") 200 people, i.e., causing them to survive while only allowing the other 400 people to die. Thus, many respondents may want to avoid program B because they do not want to gamble with the life of these 200 people that can be actively saving by choice of program A.

In turn, the choice between the programs C and D is formulated in terms of "people will die," so the reference point is 600 people alive. Therefore, many may assume the choice concerns the distribution of harms because no harm as yet has been done. Consequently, many respondents may avoid program C because they do not want to harm ("doing something for") 400 people actively and may prefer to allow a chance to decide their fate. Moreover, some authors even speculated that the descriptions of the programs C and D might suggest that people will die because of "*something lethal about the intervention itself*" (Dreisbach and Guevara 2017). Even though there is no reference to any vaccine in the original version of ADP, it is surprising that many authors speculated that programs in ADP refer to (or suggest) some vaccine whose side effects will bring about the death of most (A, C) or whose effectiveness is very uncertain (B, D).

In our experimental studies (total n=1106), we asked participants who had previously seen all four programs simultaneously to evaluate them one by one (Żuradzki et al. 2022). Specifically, we asked to respond to the following claims (on a scale 1-7; totally agree to disagree): 'This program is fair

for each of 600 people,' 'This program will harm many people,' 'This program will kill many people,' etc. Fairness seems to be not only an important reason but also risky options (B and D), which are fully egalitarian in our understanding, were indeed marked by the participants as much fairer (Figure 1).

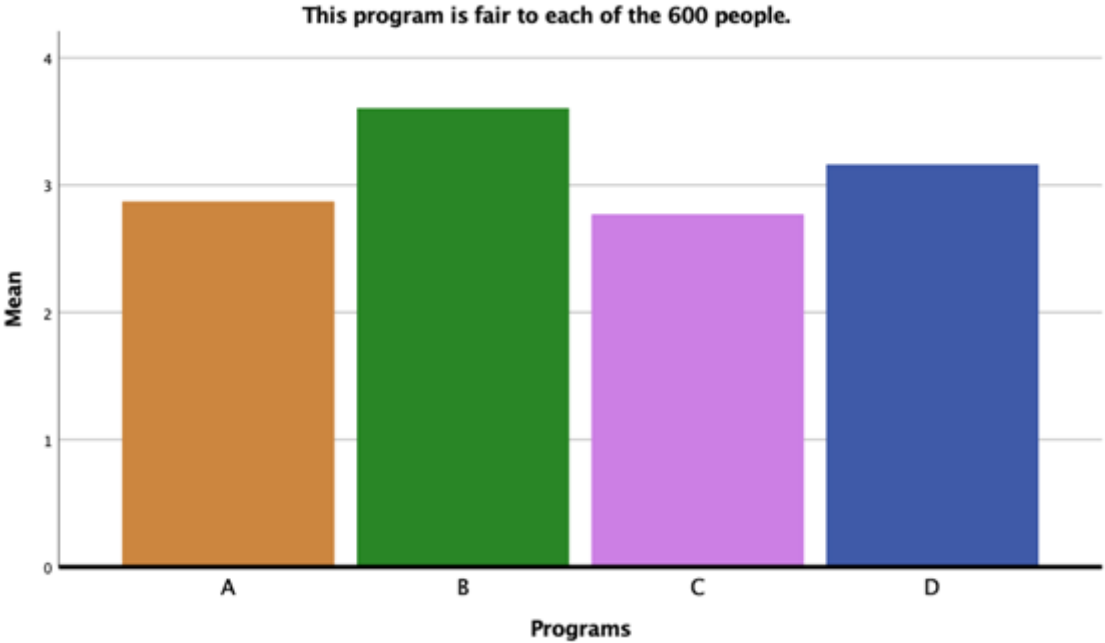


Figure 1

In this study, participants (n=164) were presented with a hypothetical scenario of a rare Asian disease approaching their country. They were told that it would kill 600 people, but four alternative programs for disease control were invented. The descriptions of the programs were presented (all at once) as in the original Asian disease paradigm. Next, participants saw descriptions of each program separately, and they were asked to evaluate them. Specifically, they were asked to respond to the following claim (on a scale 1-7; totally agree to disagree): 'This program is fair to each of the 600 people.' Program B was evaluated as the fairest to all 600 of people, in comparison to program, A ($p < .001$), program C ($p < .001$) and D ($p = .007$).

In contrast, we observed no significant differences between programs when participants evaluated whether 'This program will harm many people' or 'This program will kill many people.' In our other study, we asked participants who had previously been presented with the standard version of ADP to provide reasons why they decided on a specific program. They were presented with multiple

reasons, and they could mark up to 3. The reason 'Because this program seemed to me less harmful' was chosen relatively often (about 20-30 percent of those who chose subsequent programs marked this as one of their reason). However, we have not observed significant differences in marking this reason between the followers of different programs. Our studies show how various moral reasons may interact and counterbalance each other even in "canonical experiments" on framing effects. To sum up, our theoretical hypotheses, partially verified in the experimental studies, show that the scope of *rational framing effects* may be even broader than Bermúdez assumes in his brilliant paper.

Funding

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 805498).

References

- Bermúdez, J.J. (2021). *Frame It Again: New Tools for Rational Decision-Making*. Cambridge University Press.
- Bermúdez, J.J. (2022). Rational Framing Effects: A Multidisciplinary Case. *Behavioral and Brain Sciences*. DOI: <https://doi.org/10.1017/S0140525X2200005X>
- Dreisbach, S, Guevara, D. (2017). The Asian Disease Problem and the Ethical Implications of Prospect Theory. *Noûs* 53 (3):613-638.
- Segall, S. (2016). *Why Inequality Matters*. Cambridge University Press.
- Tversky, A., Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science* 211 (4481):453-458.
- Woollard, F. (2015). *Doing and Allowing Harm*. Oxford University Press.
- Żuradzki, T., Szwed. P., Maj M. (2022). A Risky Choice Framing Effect, Doing/Allowing, and Fairness in Uncertain Prospects (manuscript). <https://osf.io/mj3yp>