Cognition 244 (2024) 105599

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

The deflationary model of harm and moral wrongdoing: A rejoinder to Royzman & Borislow

Miklós Kürthy [a,*], Paulo Sousa [b]

[a] Department of Philosophy, University of Sheffield, Sheffield, UK
[b] Institute of Cognition and Culture, Queen's University Belfast, UK

ARTICLE INFO

Keywords
Harm
Injustice
Wrongdoing
Moral judgments
Social cognition

ABSTRACT

With a series of studies, Royzman and Borislow (2022) purport to show that extant models about the conditions under which harmful actions are deemed morally wrong do not have explanatory power—for any proposed condition, various harmful actions meet the condition but are not deemed immoral. And they reach the following conclusion: judgments of moral wrongdoing in the context of harmful actions (or judgments of moral wrongdoing more generally) are not reducible to an explanatory template. However, they did not address the main claim of the deflationary model of harm and moral wrongdoing, which is that intuitions of injustice connect harmful actions to judgments of moral wrongdoing (Sousa & Piazza, 2014). Our first study adjusts Royzman and Borislow' design to include a measure of perceived injustice, while our second elaborates their design to manipulate perceived injustice. The results undermine their conclusion and support the deflationary model, which we further refine here in light of the results of Royzman and Borislow's studies and ours.

1. Introduction

Royzman and Borislow (2022) address the puzzle of wrongless harms, namely, the fact that many actions that cause pain or suffering are not judged to be morally wrong. They do so by discussing current models about the conditions under which harmful actions are deemed morally wrong. Based on many studies, they claim that these models do not have explanatory power: "for any general pattern that is supposed to link perceptions of harm and wrongdoing, there seem to be numerous cases that match the pattern quite well but are not viewed as immoral" (Royzman & Borislow, 2022, p. 3). And they conclude more generally that "our judgments of moral wrongdoing are far too capricious, complex, conflicting, and context-dependent to be reduced to a template (or one unifying appraisal or a conditional tie)" (ibid. p. 11). One of the models discussed is the deflationary view of harm and moral wrongdoing proposed by Sousa, Piazza and colleagues (henceforth, "DfM"; see, e.g., Piazza, Sousa, Rottman, & Syropoulos, 2019; Sousa & Piazza, 2014; Sousa, Allard, Piazza, & Goodwin, 2021).[1] In this article, we question Royzman and Borislow's conclusion, showing, with two studies, that DfM is consistent with results utilising their design and that it does have explanatory power. In the remainder of this introduction, we explicate our disagreement and the rationale for our studies.

Royzman and Borislow interpret DfM as the hypothesis that the main condition that leads to the judgment that a harmful action is morally wrong is the perception that the harmful action is motivated by a selfish reason, with "selfishness" defined in terms of the agent putting their interest first, rather than that of the person being harmed. They presented participants with scenarios where a protagonist acts selfishly in this sense—e.g., a boxer hitting the opponent to win a competition (a case of physical harm) or a woman breaking up with her boyfriend to have more time for her career (a case of emotional harm). In all scenarios, with the question phrased literally in terms of whether the protagonist put their interest first, the participants acknowledged that the protagonist put their interest first while denying that the harmful action was morally wrong. Hence, Rozyman and Borislow's above conclusion.

Royzman and Borislow's interpretation of DfM does not address DfM's primary claim, which is not about selfishness but about injustice: the perception of injustice is what leads to the judgment that a harmful action is morally wrong (Sousa et al., 2021; Sousa & Piazza, 2014). Since they did not include a direct measure of perceived injustice in their

* Corresponding author.
E-mail address: m.kurthy@sheffield.ac.uk (M. Kürthy).

[1] Royzman & Borislow's main target is the Dyadic Model proposed by Gray, Schein and colleagues (e.g., Gray, Young, & Waytz, 2012; Schein, Goranson, & Gray, 2015; Schein & Gray, 2018), and they treat DfM simply as a fortified version of the Dyadic Model without using the expression "deflationary model" to refer to it. We will not consider the Dyadic Model here for we have discussed it elsewhere (see Piazza et al., 2019).

https://doi.org/10.1016/j.cognition.2023.105599
Received 13 March 2023; Received in revised form 11 August 2023; Accepted 13 August 2023
Available online 22 January 2024

studies, their results are not relevant in terms of evaluating DfM. DfM does make a secondary claim about selfishness (e.g., Piazza et al., 2019, p. 904; Sousa & Piazza, 2014, pp. 104–105), and we thank Rozyman and Borislow for making us reflect more on this topic and thereby move in a more precise direction. To clarify and elaborate, our view is that there exists another ordinary concept of selfishness related to perceptions of injustice: *a person putting their interest first over that of others in a way that violates the fair balance of interests* rather than simply *putting their interest first over those of others*.[2] (Note that we use the words "(in)justice" and "(un)fairness" interchangeably.)

Study 1 replicates Royzman and Borislow's design including a direct measure of perceived injustice, a follow-up measure to probe ordinary meanings of "selfishness", and qualitative measures concerning injustice and wrongdoing. We predicted that the results would be consistent with DfM's primary claim and would indicate that there is a sense of "selfishness" at play other than putting one's interest first. Study 2 elaborates Royzman and Borislow's scenarios to manipulate perceptions of injustice and probe whether they influence judgments of moral wrongdoing concerning harm, as per DfM's primary claim. We predicted that they would, confirming that DfM has explanatory power.[3]

2. Study 1

2.1. Participants

Participants were 301 English native speakers living in the US (154 male, 144 female, 4 other, $M_{age} = 36.52$, $SD = 13.44$, range: 19–79), each responding to two scenarios. Participants were recruited via Prolific and paid £1.50 for approximately 9 min of their time.

2.2. Design, materials, and procedures

This study was based on the design used by Royzman and Borislow's studies 1–3. There were three scenarios involving emotional harm (*Dating*, *Chess*, *Fired*) and three involving physical harm (*Mask*, *Boxing*, *Trolley*). The *Dating* scenario was as follows (see Appendix A for all the scenarios).

> Jen and Greg are classmates and have been dating for weeks. Greg has feelings for Jen and Jen really likes Greg, but, upon further reflection, Jen comes to realize that she has so much work in the coming semester that she will just have no time for this type of commitment. She could take fewer credits, but this will lower her chances of getting the summer internship she has been long planning for. She knows Greg will be depressed once he hears the news, but she also knows that this is what's best for her. So she says goodbye to Greg and puts her dating on hold.

Each participant was randomly assigned to one physical-harm scenario and one emotional-harm scenario. After reading each scenario, the participant was presented with a series of questions. The first three questions concerned moral wrongdoing. Participants were asked whether the protagonist's action was morally wrong (*moral wrongdoing probe*), the extent to which they were sure about their answer (in a sliding scale anchored as "0" for "not at all sure" and "100" for "completely sure"—*confidence probe*), and why they thought the action was wrong or otherwise, depending on their answer to the moral wrongness probe (*moral wrongdoing follow-up probe*).

Then, participants responded to four yes/no probes presented in random order (for the details, see Appendix A). Three of the probes were comprehension probes asking whether (a) the protagonist's act was intentional (*intentionality probe*), (b) the protagonist foresaw the harm would befall the other person (*foreseeable harm probe*), and (c) the protagonist put his or her interests first, rather than that of the other person (*self-interest probe*). The fourth probe (*injustice probe*) was phrased in two ways: half of the questions asked whether the protagonist acted unjustly, while the other half asked whether he/she acted unfairly. Two of these four probes had a follow-up question. Participants answering "yes" to the *self-interest probe* were also asked whether, by choosing "yes", they meant that the protagonist "was being selfish" or not (*selfishness follow-up probe*). Whether the participant answered "yes" or "no" to the injustice probe, they were further asked to explain the reasoning behind their answer (*injustice follow-up probe*). Finally, participants were asked demographic questions concerning age, gender, and religion.

2.3. Results

2.3.1. Comprehension checks and "unfair" vs. "unjust" versions of the injustice probe

Only 57 (out of 1806) responses indicated a misunderstanding of the scenario (13 in the intentionality probe, 33 in the foreseeable harm probe, and 11 in the self-interest probe). We included all responses in the following analyses. There was no significant difference between the "unjust" and the "unfair" versions of the *injustice probe*, except for the Dating scenario (*unfair* condition: 8 "unfair", 30 "not unfair"; *unjust* condition: 3 "unjust", 61 "not unjust"; Fisher's exact test, $p = .018$, $\varphi = 0.255$). We collapsed these conditions in the following analyses.

2.3.2. Two senses of "selfishness"

While the great majority of participants answered that the protagonist "put their interests first" (591 of 602), these participants were divided in their answers to the *selfishness follow-up probe* (see Table 1). Furthermore, of the participants who chose the selfish option, 110 (41%) chose the injustice option, while 158 (59%) chose the no injustice option and, more importantly, of the participants who chose the not selfish option, 309 (96%) chose the no injustice option.

2.3.3. Injustice and moral wrongdoing

Combined responses to the *injustice* and *moral wrongdoing probes* in each scenario are presented in Fig. 1. The most frequent combinations were "not unjust/not morally wrong" and "unjust/morally wrong". Accordingly, there was a strong, significant correlation between perceived injustice and moral wrongdoing in each scenario—*Dating*: $\varphi = 0.418$, $p < .001$; *Chess*: $\varphi = 0.762$, $p < .001$; *Fired*: $\varphi = 0.733$, $p < .001$; *Boxing*: $\varphi = 0.847$, $p < .001$, *Mask*: $\varphi = 0.692$, $p < .001$; *Trolley*: $\varphi = 0.540$, $p < .001$.

We ran multiple regressions within each scenario using confidence-weighted moral wrongdoing responses as an interval outcome

Table 1
Responses to the selfishness follow-up probe within each scenario.

Scenario	Selfish	Not selfish	*N*
Dating	29	68	97
Chess	19	82	101
Fired	58	39	97
Boxing	37	60	97
Mask	48	51	99
Trolley	77	23	100
Total	268	323	591

[2] In a pilot study ($N = 126$), where we asked participants directly whether the protagonists of Royzman and Borislow's scenarios (e.g., the boxer or the girlfriend) were being selfish, many participants (57 of 126) denied that they were selfish, sometimes suggesting that there were not selfish because there was no injustice involved.

[3] The studies received ethical approval from ethics committee of the School of History, Anthropology, Philosophy, and Politics, Queen's University Belfast. Both studies were pre-registered (Study 1: https://osf.io/fpg9k, Study 2: https://osf.io/9emc3). Both datasets are openly available and can be found at the pre-registration website: https://osf.io/xut3d.

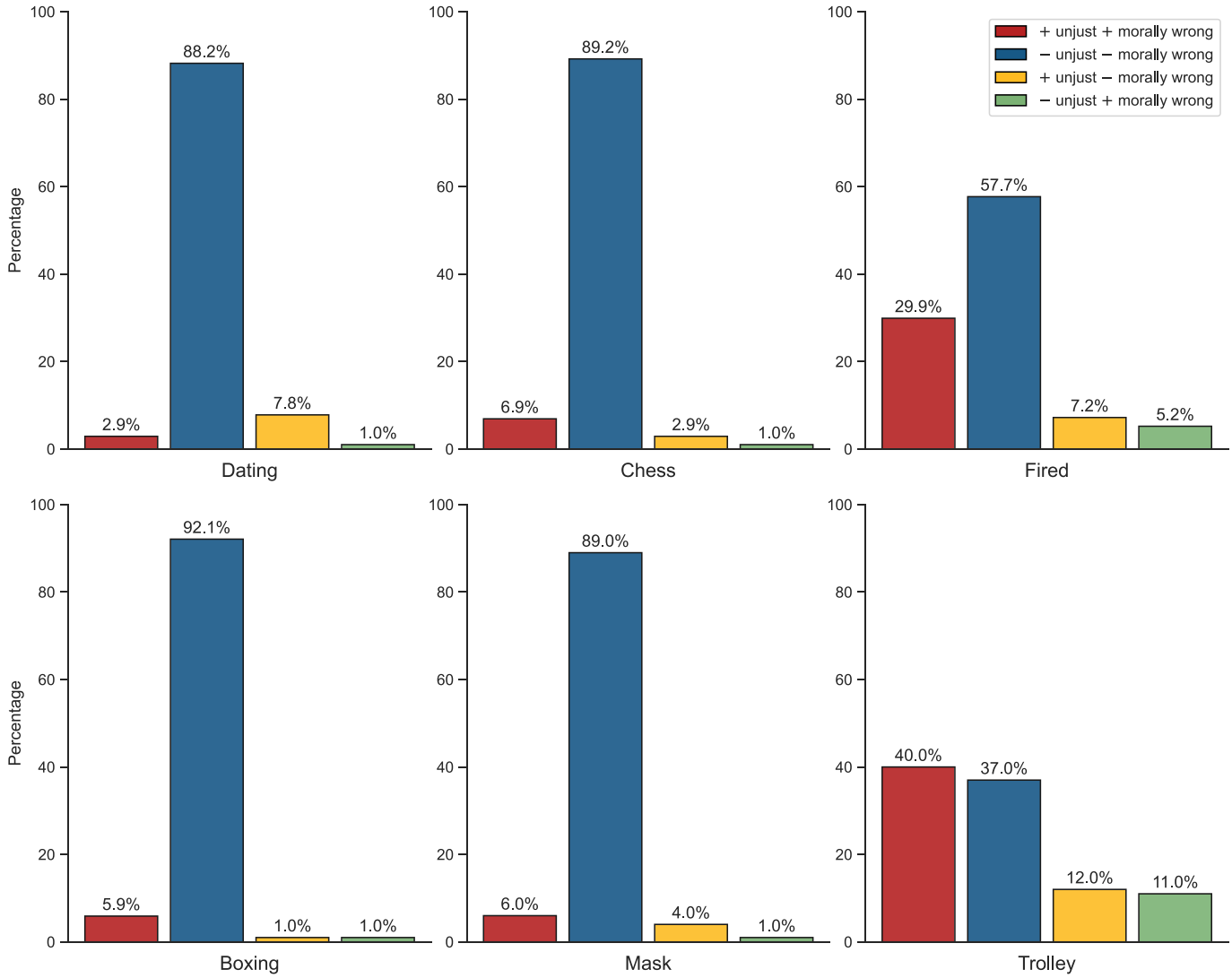

Fig. 1. Perceptions of injustice and judgments of moral wrongdoing per condition in each scenario of Study 1.

measure (see Royzman & Borislow, 2022), and intentionality, self-interest, foreseeable harm, injustice, age, and gender as predictors. Injustice was a significant predictor in all scenarios (all *ps* $< .001$). Foreseeable harm was significant in *Dating* ($p = .022$) and *Chess* ($p < .001$), and age was significant in *Fired* ($p = .013$) and *Boxing* ($p = .011$).

2.4. Discussion

The results confirm our predictions. Concerning concepts of selfishness, the fact that participants were divided in their answers to the *selfishness follow-up probe* indicates that there were two ordinary senses of "selfishness" at play. Although participants could have chosen the selfish option by virtue of applying either concept (i.e., both concepts include *a person putting their interest first over those of others*), they could have chosen the not selfish option only by virtue of applying the second concept (i.e., only the second concept incorporates *in a way that violates the fair balance of interests*, the negation of which pre-empts the perception of selfishness in this sense), which is consistent with the fact that 96% of those who chose the no injustice option chose the not selfish option.[4]

The results are also consistent with DfM's primary claim, as indicated by the phi correlations and multiple regressions. However, two limitations remain. First, since the design is merely correlational, one cannot claim that perceptions of injustice causally contribute to judgments of moral wrongdoing. Second, most participants did not see any injustice in any of the scenarios, indicating that these scenarios alone are not ideal for testing DfM's primary claim—if one wants to test whether *X* causes *Y*, it is important to check situations where *X* occurs to see whether *Y* does not occur.

[4] We are not claiming that there aren't alternative hypotheses consistent with this pattern of responses—one could invoke here an opposition between a thin (i.e., non-evaluative) concept of selfishness versus a thick (i.e., evaluative) concept of selfishness (for the literature on thin vs. thick ethical concepts, see Väyrynen, 2021). And we are not claiming that there aren't other ordinary concepts of selfishness—there may be a utilitarian-like concept of selfishness related to cases of a person not doing what, from the point of view of justice, is supererogatory. But to pursue the discussion of these points here would move us far away from the aim of this paper.

Table 2
The conditions of the *Fired* scenario. Differences between conditions appear in italics.

No injustice condition	Injustice condition
Mark is running a startup while finishing up his degree in business. *He is trying to be successful* and has five full-time employees on a one-year contract renewable at his discretion, including Georgie and Pitt, *who are the least seasoned employees*. One day, after talking things over with his financial advisor, Mark realizes that if he (a) doesn't renew the contracts of Georgie and Pitt once their contracts expire, and (b) invests in new software, *he can finally become profitable*. He knows that Georgie and Pitt will be distressed by the news since they are enjoying the work and it will take them awhile to find similar jobs. But he also knows that if he acts on this plan *he will become profitable*. So Mark goes ahead with the plan.	Mark is running a startup while finishing up his degree in business. *He has been extremely successful* and has five full-time employees on a one-year contract renewable at his discretion, including Georgie and Pitt, *who have been working exceptionally hard and have played an important role in the success of the company*. One day, after talking things over with his financial advisor, Mark realizes that if he (a) doesn't renew the contracts of Georgie and Pitt once their contracts expire, and (b) invests in new software, *he can become just a little more profitable*. He knows that Georgie and Pitt will be distressed by the news since they are enjoying the work and that it will take them awhile to find similar jobs. But he also knows that if he acts on this plan *he will be a little more profitable*. So Mark goes ahead with the plan.

Study 2 corrects these limitations. We manipulated injustice by designing two variants of four of Royzman and Borislow's scenarios, one boosting perceived injustice (injustice condition), another boosting its absence (no injustice condition). We modified their scenarios to create two conditions by taking into account both our theoretical perspective on injustice and participants' qualitative answers to the injustice follow-up probe of Study 1 (e.g., if participants said there was no injustice because of *A*, we emphasised *A* in the scenario of the no injustice condition, and if participants said that there was injustice because of *B*, we emphasised *B* in the scenario of the injustice condition).

Our theoretical perspective aligns with the mutualist-contractualist theory of justice/fairness proposed by Baumard, André, and colleagues (André, Debove, Fitouchi, & Baumard, 2022; Baumard, 2016), although judgments of culpability play a much more significant role in our perspective, and we depart from the type of moral monism they advocate.[5] On this view, the evolutionary function of our intuitive sense of justice is to regulate mutually beneficial social interactions and identify reliable partners for such interactions. In terms of proximal mechanisms involved, one may say, in a nutshell, that the intuitive sense of justice expects social interactions to obey a parameter of impartiality calibrated by a parameter of proportionality qua desert. Accordingly, to violate our intuitive sense of justice or a fair balance of interests is to be partial by not considering what people deserve. Participants' qualitative answers were consistent with this perspective.

In *Dating*, we created the injustice condition by including that Jen terminates the relationship without giving any explanation even though Greg has always been considerate to her, giving the sense that Jen is being unjust in the way she terminates the relationship because Greg does not deserve such treatment. In *Fired*, which features a boss letting his employees go, the injustice was created by including that the employees have substantially contributed to the success of the company and that the boss terminates their contracts for minimal gain, giving the sense that the boss is unjust in firing them because he neglects the proportional distribution of the burdens and benefits of cooperation. In *Boxing*, the injustice was created by including that one boxer causes harm by an illegal punch, giving the sense that the boxer thus attains an undeserved victory. Finally, in *Chess*, the injustice was created by including an experienced and dishonest player tricking an unexperienced and naïve one to play for an easy win, giving a sense of injustice due to the disproportionate advantage obtained by manipulation.

3. Study 2

3.1. Participants

Participants were 888 English native speakers living in the US (446 male, 442 female; M_{age} = 39.26, *SD* = 13.5, range: 18–93), each responding to only one scenario. Participants were recruited via Prolific and were paid £0.50 for approximately 2.5 min of their time.

3.2. Design, materials, and procedures

As we indicated above, we modified the *Dating*, *Chess*, *Fired*, and *Boxing* scenarios of Study 1, creating two versions of each: one seeking to boost perceived injustice (injustice condition), the other seeking to eliminate it (no injustice condition). Table 2 shows the two versions of the *Fired* scenario (see Appendix B for the two versions of each scenario).

Each participant was randomly assigned to one condition of a scenario (e.g., the no injustice condition of the *Fired* scenario). After reading the scenario, participants were presented with three yes/no probes: a *moral wrongdoing probe* (e.g., whether Mark's treatment of Georgie and Pitt was morally wrong), an *injustice probe* (e.g., whether Mark's treatment of Georgie and Pitt was unjust), and a *foreseeable harm probe* qua comprehension probe (e.g., whether Mark could foresee that not renewing Georgie and Pitt's contracts would make them feel distressed). The *moral wrongdoing probe* was always presented first, while *the injustice probe* and *the foreseeable harm* probe were presented in random order (see Appendix B for the probes of each scenario). Finally, participants were asked demographic questions regarding age and gender.

3.3. Results

3.3.1. Comprehension and manipulation checks

The overwhelming majority of participants (863 out of 888 or 97.2%) indicated that the protagonist foresaw the harm at stake. We included all participants in the following analyses. Indicating that our manipulation worked, the number of "yes" responses to the *injustice probe* was significantly higher in the injustice conditions than in the no injustice conditions—*Dating*: $\chi^2(1,222) = 89.26$, $p < .001$, $\varphi = 0.634$; *Chess*: $\chi^2(1, 225) = 129.51$, $p < .001$, $\varphi = 0.759$; *Fired*: $\chi^2(1, 217) = 20.93$, $p < .001$, $\varphi = 0.311$; *Boxing*: $\chi^2(1, 224) = 172.01$, $p < .001$, $\varphi = 0.876$.

3.3.2. Injustice and moral wrongdoing

The number of "yes" responses to the *moral wrongdoing probe* was significantly higher in the injustice condition than in the no injustice condition (see Fig. 2)— *Dating*: $\chi^2(1, 222) = 87.15$, $p < .001$, $\varphi = 0.627$; *Chess*: $\chi^2(1, 225) = 147.25$, $p < .001$, $\varphi = 0.809$; *Fired*: $\chi^2(1, 217) = 18.29$, $p < .001$, $\varphi = 0.290$; *Boxing*: $\chi^2(1, 224) = 159.1$, $p < .001$, $\varphi = 0.843$.

Combined responses to the *injustice* and *moral wrongdoing probes* in each scenario and injustice condition are displayed in Fig. 3. The most frequent combinations were "unjust/morally wrong" and "not unjust/not morally wrong". Accordingly, the phi correlations between the injustice and moral wrongdoing probes were all strong and significant across the conditions of each scenario—*Dating*: $\varphi = 0.831$, $p < .001$; *Chess*: $\varphi = 0.892$, $p < .001$; *Fired*: $\varphi = 0.820$, $p < .001$; *Boxing*: $\varphi = 0.893$, $p < .001$.

We performed logistic regressions with injustice condition (injustice versus no injustice) and perceived injustice (i.e., response to the *injustice probe*) as predictors and perceived moral wrongdoing (i.e., response to the *moral wrongdoing probe*) as the outcome variable. In all models,

[5] These two points will become apparent later—see General Discussion and Conclusion, respectively.

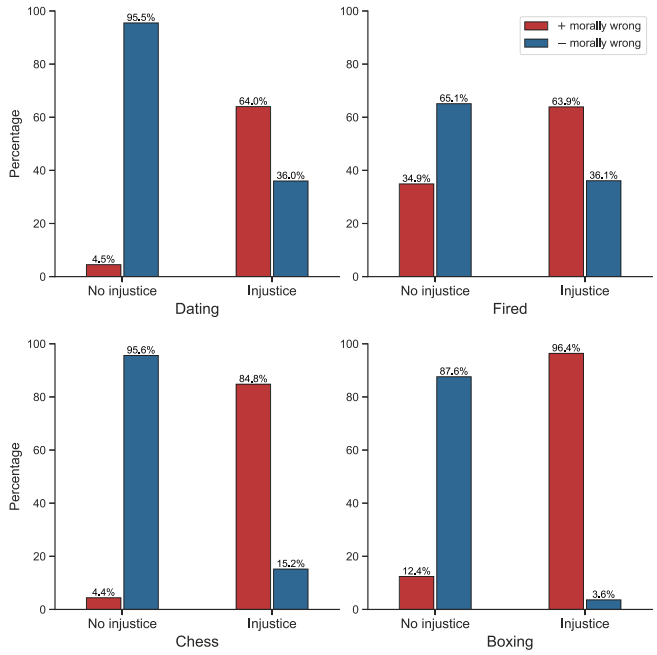

Fig. 2. Judgments of moral wrongdoing per condition in each scenario of Study 2.

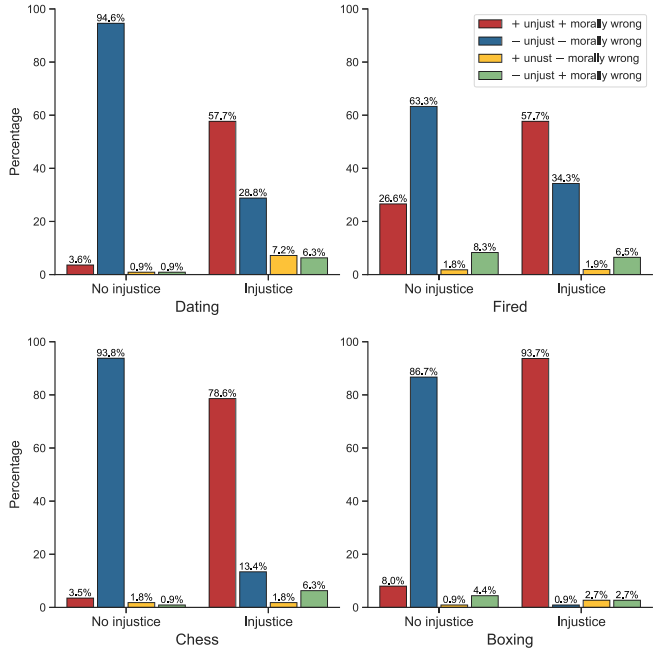

Fig. 3. Perceptions of injustice and judgments of moral wrongdoing per condition in each scenario of Study 2.

perceived injustice was significant and a stronger predictor than injustice condition. Injustice condition was significant in all but one scenario type, namely, *Fired* (see Table 3). In other words, perceived injustice was a more reliable predictor than injustice condition.

3.4. General discussion

Study 2 supports DfM's primary claim. The correlations and regressions indicated again that perceived injustice is associated with judgments of moral wrongdoing concerning harm. Since manipulating perceived injustice affected these judgments, one can interpret the association as causal. Together, our studies show that Rozyman and Borislow's conclusion is unwarranted: when perceived injustice is taken into account, judgments of moral wrongdoing concerning harm are far less capricious than they envision.

One may claim that the responses "unjust & not morally wrong" and "not unjust & morally wrong" contradict DfM. While the import of these minority responses may be limited (i.e., they may include random errors), we propose an explanation for "unjust & not morally wrong" where it was most common (i.e., *Trolley*, 12%; *Dating/Study 1*, 8%; *Dating/Injustice condition*, 7%), which requires that DfM be refined.

DfM's primary claim has been phrased as "the perception that a harmful action involves injustice leads to a judgment of moral wrongdoing". This phrasing is vague in that the expression "a harmful action involves injustice" may convey that the person's harmful action causes an unjust situation or outcome (i.e., causes unjust harm) *and/or* that the person is being unjust in causing harm. And it entails that the perception that a person's harmful action causes unjust harm alone (i.e., without the perception that the person is being unjust in causing harm) leads to a judgment of moral wrongdoing. However, we hypothesise that the concept of moral wrongdoing implies not only an evaluation of a person's action but also an evaluation of the person acting. Thus, to avoid the latter entailment, we refine DfM's primary claim as *the perception that a person is being unjust in causing harm leads to a judgment of moral wrongdoing*. Purely accidental harmful actions illustrate our point. These actions are perceived to cause unjust harm (assuming that the victim is seen as not deserving to be harmed) without the person being judged to be culpable for the causation of harm and, therefore, without the person being perceived as unjust in causing the harm.[6] Since, concerning these actions, the evaluation of the person causing harm is blocked by a judgment of no culpability, DfM's refined formulation entails a judgment of no moral wrongdoing.

In our studies, the phrasings of the *injustice probe* reflect the above vagueness, leaving open the possibility that the injustice option be chosen simply to convey that the harmful action/treatment caused unjust harm. Purely accidental harmful actions are not the only case where the evaluation of the person acting is blocked. This may happen with other harmful actions if, all things considered, the reason for acting is deemed to justify the person/action. In the *Trolley* scenario, saving one's life may constitute such a reason, leading to a judgment of no moral wrongdoing. Moreover, this scenario involves a particularly unjust outcome (an undeserved death), which may have led some of the same participants to choose the injustice option. In the *Dating* scenarios, there is a common assumption that staying in a dating relationship is discretionary. In Study 1, for most participants, this understanding completely justifies the voluntary termination of the relationship, leading to a judgment of no moral wrongdoing. In the injustice condition of Study 2, there is also the fact that the way the relationship is terminated violates our sense of justice. Here, some participants may have thought that the

[6] By pure accident, we mean that there is not even the perception of negligence (i.e., that the person did not foresee the harm but should have foreseen it). Although we do not have space to discuss all our assumptions here, let's lay them out. We assume (i) that considerations of whether an action is intentional or not are relevant to judgments of moral wrongdoing only as a component of judgments of culpability, (ii) that the most relevant distinction for the boundary of the concept of moral wrongdoing is that between two types of non-intentional causation of harm: causing harm by a pure accident, which implies no culpability, and causing harm by negligence, which implies some culpability, (iii) and that the perception of a person being unjust in causing harm implies that the person has at least some degree of culpability for the harm. For our view on folk concepts of intentional action, see Sousa & Holbrook, 2010; Sousa, Holbrook, & Swiney, 2015; Sousa & Lavery, 2023. For our view on judgments of culpability, see Sousa, 2009b; Sousa & Swiney, 2016; Sousa & Allard, 2018; Sousa & Lavery, 2023 (see also Zimmerman, 1988, 2011).

Table 3
Injustice condition and perceived injustice as predictors of judgments of moral wrongdoing.

Scenario	Predictor	Odds ratio	95% CI	*p*-value
Dating	Injustice condition	7.95	(2.33, 27.11)	.001
	Perceived injustice	56.00	(19.68, 159.36)	< .001
Chess	Injustice condition	34.14	(7.9, 147.65)	< .001
	Perceived injustice	122.85	(29.55, 510.7)	< .001
Fired	Injustice condition	1.58	(0.62, 4.05)	.34
	Perceived injustice	135.95	(43.49, 425.02)	< .001
Boxing	Injustice condition	14.54	(3.28, 64.56)	< .001
	Perceived injustice	60.22	(14.71, 246.45)	< .001

discretionary assumption is overriding, or they may have misunderstood the *moral wrongdoing probe* as being about the action of terminating the relationship (rather than being about *the way that* the relationship was terminated). Either way, this would lead to a judgment of no moral wrongdoing. Moreover, in both studies, the same participants could still have the perception that the end of the relationship caused an unjust situation, given that the scenarios may have suggested that the partner is a good, loving person, or at any rate that he did not deserve to suffer, and choose the injustice option. In sum, if participants' "unjust & not morally wrong" responses in these scenarios convey simply that the action created unjust harm without person evaluation, they actually correspond to what the refined version of DfM predicts.

4. Conclusion

We conclude by clarifying our perspective on some broader issues and acknowledging an important limitation.

We are not advancing any kind of moral monism. DfM implies that there isn't a separate moral domain based on a concept of harm *qua* pain or suffering (or, more broadly, welfare reduction) since, according to it, the moralisation of harm depends on its perceived injustice. However, it does not imply that there aren't moral domains unrelated to intuitions of injustice, and we are open to this possibility (see Piazza & Sousa, 2023; Sousa & Piazza, 2014). Concerning intuitions of injustice, cognitive scientists have broadly distinguished between distributive, retributive/restorative, and procedural justice, and occasionally they have made even more fine-grained distinctions between injustice-related psychological mechanisms (Baumard, André, & Sperber, 2013; Bøggild & Petersen, 2016; Boyer, 2015; Cosmides, 1989; Darley & Shultz, 1990; Delton, Cosmides, Guemo, Robertson, & Tooby, 2012; Goodwin & Gromet, 2014; Hamann, Warneken, Greenberg, & Tomasello, 2011). For example, some authors distinguish perceptions of oppression/liberty from perceptions of cheating/reciprocity in terms of the psychological mechanisms involved (Haidt, 2012).[7] Although for the sake of simplification we have phrased our discussion in this paper in terms of *the* sense of (in)justice, we agree that there are various injustice-related psychological mechanisms and, by individuating moral domains in terms of the computational profile of psychological mechanisms (Carruthers, 2006), propose that there are various moral domains related to injustice.

Some researchers do not bother with specifying moral wrongdoing since their work operationalises the concept of moral wrongdoing by simply asking participants whether an action is wrong (or, equivalently, impermissible)—as if by probing a superordinate concept, one could obtain clear evidence about a subordinate one. Other researchers use words such as "morally (wrong)" or "immoral (action)" to do the job—as if English words that are quite polysemous and do not translate to many languages could say something fundamental about human normative thinking. Similarly to Royzman's highly sophisticated work on the topic (e.g., Royzman, Landy, & Goodwin, 2014; Royzman, Leeman, & Baron, 2009), we normally prefer to specify moral wrongdoing in terms of a type of normative conviction that may be operationalised by criteria such as generalizability and authority-independence à la the Turiel tradition (Turiel, 1983; see also Bartels, Bauman, Cushman, Pizarro, & McGraw, 2015), and DfM has been tested and has fared well in the context of this approach (see Berniūnas, Dranseika, & Sousa, 2016; Piazza & Sousa, 2016; Piazza, Sousa, & Holbrook, 2013; Sousa, 2009a; Sousa et al., 2021; Sousa, Holbrook, & Piazza, 2009; Sousa & Piazza, 2014). We adopted the second approach here (and both this and the first approach elsewhere—see Piazza et al., 2019) in order to establish a fruitful dialogue with the literature (i.e., here to make our results commensurable with Royzman & Borislow's results).[8] It is reassuring that DfM is supported in the context of this second approach too. Nonetheless, one limitation of our results is that they may merely indicate that the concept UNJUST HARM (or BEING UNJUST IN CAUSING HARM) is a prototypical meaning of words such as "morally (wrong)" and "immoral" when applied to actions, as suggested by some dictionaries.

Credit author statement

The authors have contributed equally to this work.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Data availability

The data that support the findings of this study are openly available at https://osf.io/xut3d/files/osfstorage.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105599.

References

André, J. B., Debove, S., Fitouchi, L., & Baumard, N. (2022). An evolutionary contractualist account of morality. *PsyArXiv*, 1–48. https://doi.org/10.31234/osf.io/2hxgu

Bartels, D., Bauman, C. W., Cushman, F. A., Pizarro, D., & McGraw, P. A. (2015). Moral judgment and decision making. In G. Keren, & G. Wu (Eds.), *Vol. 1. The Wiley Blackwell handbook of judgment and decision making* (pp. 478–515). Wiley Blackwell.

Baumard, N. (2016). *The origins of fairness: How evolution explains our moral nature*. Oxford University Press.

[7] Haidt may not accept that both psychological mechanisms here concern injustice/justice intuitions. We take these and other psychological mechanisms to concern injustice/justice intuitions because they evolved in response to problems in the evolution of human cooperation and because their concepts, as expressed by ordinary words such as "unjust/just", "unfair/fair", "desert", share important commonalities.

[8] We presume Royzman has also adopted different approaches in this respect for similar reasons.

Baumard, N., André, J. B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences, 36*(1), 59–78.
Berniūnas, R., Dranseika, V., & Sousa, P. (2016). Are there different moral domains? Evidence from Mongolia. *Asian Journal of Social Psychology, 19*, 275–282.
Bøggild, T., & Petersen, M. B. (2016). The evolved functions of procedural fairness: An adaptation for politics. In T. K. Shackelford, & R. D. Hansen (Eds.), *The evolution of morality* (pp. 247–276). Springer.
Boyer, P. (2015). How natural selection shapes conceptual structure: Human intuitions and concepts of ownership. In E. Margolis, & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 185–200). MIT Press.
Carruthers, P. (2006). *The architecture of the mind: Massive modularity and the flexibility of thought*. Oxford University Press.
Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*(3), 187–276.
Darley, J., & Shultz, T. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology, 41*, 525–556.
Delton, A. W., Cosmides, L., Guemo, M., Robertson, T. E., & Tooby, J. (2012). The psychosemantics of free riding: Dissecting the architecture of a moral concept. *Journal of Personality and Social Psychology, 102*(6), 1252–1270.
Goodwin, G. P., & Gromet, D. M. (2014). Punishment. *WIREs. Cognitive Science, 5*(5), 561–572.
Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry, 23*(2), 101–124.
Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage Books.
Hamann, K., Warneken, F., Greenberg, J. R., & Tomasello, M. (2011). Collaboration encourages equal sharing in children but not in chimpanzees. *Nature, 476*, 328–331.
Piazza, J., & Sousa, P. (2016). When injustice is at stake, moral judgements are not parochial. *Proceedings from the Royal Society of London B, 283*, 20152037.
Piazza, J., & Sousa, P. (2023). Minimal criteria for an impurity domain of morality. *Trends in Cognitive Science, 27*(6), 514–516.
Piazza, J., Sousa, P., & Holbrook, C. (2013). Authority dependence and judgments of utilitarian harm. *Cognition, 128*, 261–270.
Piazza, J., Sousa, P., Rottman, J., & Syropoulos, S. (2019). Which appraisals are foundational to moral judgment? Harm, injustice, and beyond. *Social Psychological and Personality Science, 10*(7), 903–913.
Royzman, E. B., & Borislow, S. H. (2022). The puzzle of wrongless harms: Some potential concerns for dyadic morality and related accounts. *Cognition, 220*, Article 104980.
Royzman, E. B., Landy, J. F., & Goodwin, G. P. (2014). Are good reasoners more incest friendly? Trait cognitive reflection predicts selective moralization in a sample of American adults. *Judgment and Decision making, 9*(3), 176–190.
Royzman, E. B., Leeman, R. F., & Baron, J. (2009). Unsentimental ethics: Towards a content-specific account of the moral-conventional distinction. *Cognition, 112*(1), 159–174.
Schein, C., Goranson, A., & Gray, K. (2015). The uncensored truth about morality. *The Psychologist, 28*(12), 982–985.
Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review, 22*(1), 32–70.
Sousa, P. (2009a). On testing the moral law. *Mind & Language, 24*, 209–234.
Sousa, P. (2009b). A cognitive approach to moral responsibility: The case of a failed attempt to kill. *Journal of Cognition and Culture, 9*(3–4), 171–194.
Sousa, P., & Allard, A. (2018). *Reasons concerning unintentional harm, obligations concerning intentional harm: A critique of the path model of blame*. Unpublished Manuscript.
Sousa, P., Allard, A., Piazza, J., & Goodwin, G. P. (2021). Folk moral objectivism: The case of harmful actions. *Frontiers in Psychology, 12*, 1–18.
Sousa, P., & Holbrook, C. (2010). Folk concepts of intentional action in the contexts of amoral and immoral luck. *Review of Philosophy and Psychology, 1*(3), 351–370.
Sousa, P., Holbrook, C., & Piazza, J. (2009). The morality of harm. *Cognition, 113*, 80–92.
Sousa, P., Holbrook, C., & Swiney, L. (2015). Moral asymmetries in judgments of agency withstand ludicrous causal deviance. *Frontiers in Psychology, 6*, 1380.
Sousa, P., & Lavery, G. (2023). Culpability and liability in the law of homicide: Do lay moral intuitions accord with legal distinctions? In K. Prochownik, & S. Magen (Eds.), *Advances in experimental philosophy of law* (pp. 99–132). Bloomsbury Academic.
Sousa, P., & Piazza, J. (2014). Harmful transgressions qua moral transgressions: A deflationary view. *Thinking & Reasoning, 20*(1), 99–128.
Sousa, P., & Swiney, L. (2016). Intentionality, morality, and the incest taboo in Madagascar. *Frontiers in Psychology, 7*, 494.
Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge: Cambridge University Press.
Zimmerman, M. J. (1988). *An essay on moral responsibility*. New Jersey: Rowman & Littlefield.
Zimmerman, M. J. (2011). *The immorality of punishment*. Broadview Press.