




ARTICLE

# Description invariance: a rational principle for human agents

Sarah A. Fisher 

UCL Department of Political Science, The Rubin Building, 29/31 Tavistock Square, London WC1H 9QU, UK  
Email: [sarah.a.fisher@ucl.ac.uk](mailto:sarah.a.fisher@ucl.ac.uk)

(Received 04 January 2022; revised 04 November 2022; accepted 23 November 2022; first published online 16 February 2023)

## Abstract

This article refines a foundational tenet of rational choice theory known as the principle of description invariance. Attempts to apply this principle to human agents with imperfect knowledge have paid insufficient attention to two aspects: first, agents' epistemic situations, i.e. whether and when they recognize alternative descriptions of an object to be equivalent; and second, the individuation of objects of description, i.e. whether and when objects count as the same or different. An important consequence is that many apparent 'framing effects' may not violate the principle of description invariance, and the subjects of these effects may not be irrational.

**Keywords:** Description invariance; rational choice theory; framing effects; equivalence; imperfect knowledge

## 1. Introduction

Rational choice theory makes some basic, minimal assumptions about the coherence of agents' evaluative judgements. One such assumption is captured by the principle of description invariance. In its idealized form, this principle states that perfectly rational agents with perfect knowledge will never evaluate the same thing differently just because it is described in different ways. For example, such agents will not rate the oratory of Cicero more favourably than that of Tully, or vice versa, since (i) they know that the same person is referred to by the names 'Cicero' and 'Tully' and (ii) they have a consistent view of that individual's oratory. Because the perfectly knowledgeable agent knows that the same person is referred to by the names 'Cicero' and 'Tully', and also knows everything there is to know about this person, it is impossible that the choice to refer to him by one or other name could itself convey any additional information about his oratory. Meanwhile, because the agent is perfectly rational, she cannot simultaneously hold inconsistent attitudes towards a single entity. Otherwise she would be incapable of making all sorts of ordinary

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

decisions (Broome (1993) makes a similar point). Accordingly, our idealized agent cannot rate the oratory of Cicero differently from that of Tully.<sup>1</sup>

Below is a general formulation of the description invariance principle, as it applies to agents who are perfectly rational and omniscient:

- (1). Any given object is evaluated in the same way under alternative descriptions.

Both ‘object’ and ‘description’ are to be understood broadly here. Thus, an object could be an individual (like Cicero), another kind of entity, a property, or a state of affairs. Meanwhile, a description could be a name (like ‘Cicero’ or ‘Tully’), a predicate, or a sentence.<sup>2</sup>

Unlike the ideal agents we have just been considering, humans do not have perfect knowledge. Yet it is important for theorists to be able to distinguish human behaviour that is rational from that which is irrational, and to keep that distinction separate from questions of knowledge and ignorance. Imagine, for example, a student of history who greatly admires Cicero’s oratory in the Catiline Orations but is nonplussed by many of Tully’s legal defences. The student fails to realize that Cicero *is* Tully, i.e. that the names ‘Cicero’ and ‘Tully’ refer to the same historical figure. Our student ends up rating the oratory of Cicero higher than that of Tully. Yet this seems not to be the result of a rational coherence error but rather of a lack of knowledge. The principle of description invariance should allow, then, that rational human agents can evaluate the same thing differently under different descriptions if they *fail to know* that the descriptions share a single object. That would point to a modification of our earlier principle as follows:

- (2). Any given object is evaluated in the same way under alternative descriptions unless it is not known to the agent that the descriptions share an object.

However, principle (2) cannot be accepted as it stands. Human agents are not *always* entitled to evaluate the same thing differently under different descriptions, just because they fail to know that the descriptions share an object. Our imagined student is entitled to rate the oratory of Cicero higher than that of Tully because he believes that Cicero and Tully were two different individuals

---

<sup>1</sup>Note that rational choice theory is silent on the question of precisely which kinds of entities evaluative judgements are defined over. So, in principle, we could individuate the oratory of Cicero in more fine-grained ways: as the oratory of person  $x$  insofar as we use the name ‘Cicero’ to refer to  $x$ , on one hand; or as the oratory of person  $y$  insofar as we use the name ‘Tully’ to refer to  $y$ , on the other. However, since the perfectly knowledgeable agent knows that  $x = y$ , there is no motivation for doing so here. The name used cannot affect the agent’s information about the individual or his oratory. Since there is no room for the distinct names to carry additional information, there is no justification for finer-grained individuation.

<sup>2</sup>No assumptions are made about the precise way in which a description, broadly understood, connects to its object (for example, whether via rigid or non-rigid designation). That is an orthogonal issue that need not concern us here.

with different oeuvres, Cicero's being superior to Tully's.<sup>3</sup> This kind of justification will not always exist. Consider, for example, a customer in a clothes store who tries on several dresses. The personal shopper asks her to rate the fit of each dress. The final dress is yellow. Unbeknownst to the customer, it is made by Givenchy, which she knows to be a prestigious brand. After discussing each of the other dresses, the personal shopper might ask 'And how about the fit of the final dress – the yellow one?'. Alternatively, she might ask 'And how about the fit of the final dress – the Givenchy one?'. Intuitively, the dress's actual fit is entirely unaffected by whether it is subsequently described as 'yellow' or 'Givenchy'. Whatever else is conveyed to the customer by these alternative descriptions, they do not carry any further information about the dress's fit. Therefore, insofar as the customer would rate its fit higher under the latter description than the former, this could not be given any justification. Rather, it would seem to indicate mere snobbery.

What the dress case shows is that a mere failure to know that two descriptions share an object is not a sufficient condition for rational agents to evaluate the object differently under each description. Instead, the lack of knowledge must be accompanied by a belief that the descriptions (do or could) have distinct objects with distinct properties, and that these properties differ in ways that would justify distinct evaluations. Our imagined student meets this additional criterion. His failure to realise that Cicero is Tully is relevant to his subsequent evaluations of Cicero's/Tully's oratory, since the student believes that there is one individual answering to 'Cicero' who orated the well-regarded Catiline Orations, and another individual answering to 'Tully' who orated various other less well-regarded legal defences. The student believes the distinct names refer to distinct individuals with distinct oeuvres of varying quality; in other words, for the student, the names carry distinct information about the evaluatively relevant property. Therefore, in rating Cicero's oratory higher than Tully's, he does not violate rational coherence.

In contrast, our imagined clothing customer does violate rational coherence: failing to know that the yellow dress is made by Givenchy cannot justify rating its fit lower. In general, then, alternative descriptions must carry relevantly different information about the object of evaluation, at least from the epistemic perspective of the agent with imperfect knowledge. The agent must believe that the descriptions are applicable to entities with distinct properties; and those properties must be relevant, in principle, to the evaluation task at hand. Only then can we consider an agent's lack of knowledge about the co-extension of alternative descriptions as *evaluatively relevant*.<sup>4</sup>

---

<sup>3</sup>I leave aside the question of whether the student *ought* to have found out that the names 'Cicero' and 'Tully' are coreferential. Instead, I simply take his epistemic state as given. While the culpability or non-culpability of someone's ignorance might bear on a more substantive, all-things-considered notion of rationality, it is not relevant to the minimal coherence issue we are concerned with here.

<sup>4</sup>In contrast with the case of the omniscient agent, it now becomes reasonable to individuate the objects of evaluation in a more fine-grained way. For example, it would make sense to define the evaluative judgements of our imagined student not over the oratory of Cicero *simpliciter*, but over the oratory of person *x* insofar as we use the name 'Cicero' to refer to *x*, as distinct from the oratory of person *y* insofar as we use the name 'Tully' to refer to *y*. Our agent with imperfect knowledge does *not* believe  $x = y$ .

Building these considerations into the description invariance principle requires something like the further modification below:

(3). Any given object is evaluated in the same way under alternative descriptions unless (a) it is not known to the agent that the descriptions share an object, such that (b) it is believed to be possible that the descriptions refer to distinct objects with distinct properties, and (c) those distinct properties, if they obtained, would justify different evaluations.<sup>5</sup>

Again, since there would be good grounds for evaluating individuals' oratories differently on the basis of distinct oratorial canons, principle (3) is not violated by our imagined student. In contrast, since there would be no grounds for our imagined clothing customer to evaluate the dress's fit differently on the basis of its coming from a prestigious designer, principle (3) *is* violated.

I take it that something like principle (3) is implicitly assumed in standard theories of rational choice (although I am not aware of its being spelled out so explicitly – see also Fisher and Mandel (2021) for further discussion). I will now argue that it must be refined still further. Specifically, more care is needed in relation to the characterization of agents' epistemic states (to be discussed in §2) and the individuation of objects of description (to be discussed in §3). As we will see, the proposed refinements bear importantly on the interpretation of psychological 'framing effects' (to be discussed in §4). Namely, while these are usually understood as paradigmatic violations of the principle of description invariance, and thus as evidence of human irrationality, I will argue that this is far from being established.

## 2. Knowing What is Known

Assessing agents' compliance with principles (2) or (3) already requires assessing their epistemic situation – namely, their beliefs about whether or not two descriptions share an object. This requirement is a direct consequence of the principle being tailored to agents with imperfect knowledge. Assessing omniscient agents' compliance with principle (1) only requires consideration of whether or not the two descriptions do, in fact, share an object. By assumption, omniscient agents' epistemic states already track the facts perfectly. Therefore, if two descriptions share an object, the omniscient agent knows that they do, and should evaluate the object consistently. In contrast, agents with imperfect knowledge may not.

The need for epistemic assessment ushers in various problems, however. In some cases (like the Cicero-Tully case discussed earlier) it is simple enough to establish what agents believe and, in turn, whether or not they comply with principle (3). In other cases it is more difficult, and these include some paradigm psychological 'framing effects'. Consider, for example, the case of a basketball player, who can

---

<sup>5</sup>While intuitions may vary concerning exactly which differences in properties would, in fact, justify different evaluations, the examples given in the paper are sufficiently clear-cut as to avoid this complication for now.

be described as having *made* 40% of his shots, or as having *missed* 60% (Leong *et al.* 2017). Psychological experiments confirm that the player is typically rated higher under the first ‘made’ frame than under the second ‘missed’ frame (Leong *et al.* 2017). Such shifts in people’s judgements have been reproduced again and again throughout the large literature on framing.<sup>6</sup> The findings are standardly taken as proof of humans’ systematic and pervasive irrationality. Since the alternative frames quite obviously describe the same object, it is claimed, agents violate the principle of description invariance – understood as (3) above – whenever their evaluative judgements reveal frame-sensitivity.

Against this dominant interpretation, a handful of researchers have begun to question the assumption that experimental participants really do recognize the frames to be equivalent. An alternative possibility is that they are understanding numerical quantifiers as representing *lower bounds* rather than exact or approximate quantities. For example, the sentence, ‘The player made 40% of his shots’ might be understood to mean that he made *at least* 40%, while ‘The player missed 60% of his shots’ might be understood to mean that he missed *at least* 60%. It would then be perfectly reasonable to evaluate the player more favourably under the ‘made’ frame than under the ‘missed’ frame. After all, if he made at least 40% of his shots, he could have missed less than 60%. In contrast, if he missed at least 60%, he could only have made at most 40%. In other words, his performance could be objectively better under the ‘made’ frame than under the ‘missed’ frame. First mooted at least as far back as MacDonald (1986), this proposal has recently acquired empirical support from Mandel (2014); Mandel finds a link between lower-bounded interpretations of numerical quantifiers and the emergence of framing effects.<sup>7</sup>

Agents who form lower-bounded interpretations of the numerical quantifiers used in framing study stimuli would seem not to violate the description invariance principle: from their epistemic perspective, each frame could be used to describe different players with different shooting performances. Importantly, this epistemic possibility is also evaluatively relevant. As we saw, from the epistemic perspective of such an agent, the player’s performance could be objectively better under the ‘made’ frame than under the ‘missed’ frame, justifying the observed shift in the assessment of that performance.

This is a little too hasty, though. It might be objected that the argument from lower-bounded interpretations relies on a faulty conception of the agent’s epistemic state. Such an objection might be mounted by appeal to the ‘reflection test’ put forward by Tversky and Kahneman in their discussions of framing effects. They write:

Two characterizations that the decision maker, on reflection, would view as alternative descriptions of the same problem should lead to the same choice – even without the benefit of such reflection. (Tversky and Kahneman 1986: S253)

<sup>6</sup>For a survey of the first 30 years of framing research, see Levin *et al.* (1998).

<sup>7</sup>But see Simmons and Nelson (2013) and Chick *et al.* (2016) for demonstrations that lower-bounded interpretations of quantifiers cannot explain framing effects in their entirety.

Thus, Tversky and Kahneman argue that what matters for rationality is whether or not reasoners would judge alternative frames to be equivalent *on reflection*. Kahneman (2000: xv) further elaborates on what reflection is supposed to involve, writing:

It is the decision maker who should determine, after due consideration of both problems, whether the differences between them are sufficiently consequential to justify different choices.

The idea seems to be that we would need to elicit agents' beliefs *after they have duly considered each of the descriptions*. Note that this necessarily involves a change in context, from one in which agents just consider one frame and evaluate the entity it describes, to one in which they consider both frames and judge whether the differences between them could justify different evaluations. According to the proponents of the reflection test, it is the judgements that emerge from the second kind of context that establish whether or not the description invariance principle has been violated in the first kind of context.

What should we say, then, about agents who form lower-bounded interpretations of numerical quantifiers? If we adopt the reflection test, this will depend on the beliefs they have (or would have) after reflecting on both frames. Returning to the case of the basketball player, agents who initially formed a lower-bounded interpretation of the percentage figure in one or the other frame might nevertheless conclude, after due consideration of both frames, that each is being used to describe exactly the same performance (and, accordingly, that the percentages denote exact or approximate quantities rather than lower-bounded ones). In light of this revised interpretation, let's assume that the agents would no longer consider the difference in framing to justify different evaluations of the player.<sup>8</sup> Any prior tendency to rate the player more favourably under the 'made' frame than under the 'missed' frame would now be deemed irrational on that basis. Indeed, several studies show that people typically do avoid making distinct judgements once they have had the opportunity to consider both frames (Frisch 1993; Kühberger 1995; Stanovich and West 1998).

In fact, though, I do not believe we should accept the reflection test – at least not without further evidence of its utility. Presenting multiple frames together, for consideration of any evaluatively relevant differences between them, is an importantly different task from presenting each frame separately, for straightforward evaluation of the object being described. It is possible that presentation in different contexts (joint vs. separate) and for different purposes (higher-order vs. first-order judgements) could affect the interpretation of the stimuli themselves. So, for example, when making a higher-order judgement in a joint presentation context, it might turn out to be more natural to adopt exact or approximate interpretations of numerical quantifiers.<sup>9</sup> If this possibility

<sup>8</sup>Although see §3 for a reason why they might still do so.

<sup>9</sup>That could help make sense of the experimenters' intentions in presenting frames that can be used to describe the same thing but tend to produce distinct reactions in us. In contrast, the same interpretative pressure is absent when one is presented only with a single frame and asked to evaluate the object it

proved correct, the reflection test would be shown to be illegitimate, since it assumes that the stimuli are ultimately to be interpreted in the first way (i.e. as involving exact or approximate interpretations of numerical quantifiers). Instead, though, in some other contexts it could be perfectly reasonable to derive the second interpretation (i.e. as involving lower-bounded interpretations of numerical quantifiers).

One might object again at this point that if experimental participants' interpretations were changing across contexts, they would surely recognise and report this. There is little evidence from the framing literature that they do so, which may lead us to think that their interpretation does not in fact change. In response to this objection, I do not believe it is always realistic to expect agents with imperfect knowledge to have the requisite access to their interpretations to recognise and report them – perhaps they are simply unable to do so. In that case, their responses in different contexts could end up looking mutually inconsistent – even to the agents themselves – while in fact being arrived at quite reasonably, on the basis of relevantly different implicit interpretations.

Of course, the proposal sketched in this section is speculative and in need of empirical confirmation. The important point for now is as follows: one cannot simply *assume* that agents' interpretations of linguistic expressions in one context have normative force over their interpretations in other contexts. Instead, agents might reasonably form different interpretations in different contexts, which in turn justify their different evaluative judgements.

The general lesson to draw from the preceding discussion is that the knowledge criterion in the principle of description invariance must be relativized to agents' contextual interpretations of the descriptions in question. We can make this explicit in a further refinement of the principle, as follows:

- (4). Any given object is evaluated in the same way under alternative descriptions unless (a) it is not known to the agent that the descriptions (as interpreted by the agent in the context of evaluation) share an object, such that (b) it is believed to be possible that the descriptions refer to distinct objects with distinct properties, and (c) those distinct properties, if they obtained, would justify different evaluations.

The need for the parenthetical clause in (4) reflects the fact that assessing agents' compliance with the principle depends on a more nuanced and context-sensitive assessment of their epistemic states than is standardly assumed. In particular, we need to isolate the interpretations and beliefs they form in the particular task environment at issue.

### 3. Individuating Objects

A further set of problems concerns the way in which we individuate objects of description. Let's return to the case of the basketball player and assume for

---

describes. Perhaps, then, it is more natural in that context to interpret the numerical quantifier as denoting a lower bound.



simplicity that the percentages are understood as exact, so that the player made *exactly* 40% and missed *exactly* 60% of his shots. Experimental psychologists have observed that the alternative frames are still unlikely to be entirely informationally equivalent (Sher and McKenzie 2006, 2008, 2011). This is because language users' framing choices are sensitive to features of the context. When the player being described made a relatively large proportion of shots (compared, say, with a typical player) speakers tend to opt for the 'made' frame; conversely, when the same absolute performance counts as making a relatively small proportion of shots, they err towards using the 'missed' frame (Leong *et al.* 2017). Their audiences are, in turn, sensitive to the association between frame and context: when presented with the 'made' frame, audiences have a greater tendency to infer that the player made a relatively large proportion of shots than when the 'missed' frame is used, and vice versa (Leong *et al.* 2017). This can help explain why the player is evaluated more favourably under the 'made' frame than under the 'missed' frame. Moreover, it helps *justify* these frame-sensitive judgements. After all, the frames convey distinct information about the player's *relative* performance, which supplements the common information each frame carries concerning his *absolute* performance. Information about the player's relative performance is, in turn, relevant to evaluating him (at least where hearers lack overriding knowledge of what counts as good or bad performance in absolute terms). Other things being equal, a player whose shooting rate is understood to be relatively high (compared, say, with the typical player) should be evaluated more favourably than a player whose shooting rate is understood to be relatively low. In general, it is perfectly reasonable for agents to make use of linguistic cues concerning *relative* performance (and not just *absolute* performance) when assessing a player.<sup>10</sup>

Note here that there are two different ways we might construe the inference of the contextual information from a speaker's choice of frame. On one construal, the additional piece of information is inferred with *certainty*. Thus, someone who assumes that the 'made' frame is only ever used to describe relatively good shooting performance, while the 'missed' frame is only ever used to describe relatively bad shooting performance, will infer with certainty the information about the player's relative shooting performance. Alternatively (and more plausibly) someone who assumes that the 'made' frame is merely *more likely* to describe relatively good shooting performance, while the 'missed' frame is more likely to describe relatively bad shooting performance, will assign some higher-than-chance probability to that information. On the probabilistic construal, the use of the 'made' frame is taken to raise the probability that the player's shooting performance is relatively good – and vice versa for the 'missed' frame – without this being treated as guaranteed.

Note also how the justification for framing effects in this instance could not be applied to the dress example discussed earlier. In that example, the alternative descriptions, 'yellow' and 'Givenchy' do not plausibly convey any additional

<sup>10</sup>It should be noted, of course, that such cues can be defeated by a number of other factors on any specific occasion. For example, the speaker might be known to be insincere or unreliable. I set these issues aside for now in order to draw the general lesson.



contextual information concerning the dress's fit, either in absolute terms or as compared with the other dresses the customer tried on. Any *prima facie* information about fit that might have been conveyed by the brand name is overridden by the information the customer acquires from actually trying on the dress.

Sher and McKenzie's insights concerning the contextual information conveyed by frames must also be integrated into our formulation of the description invariance principle. We want to say that agents can be rational to infer and use this additional information, wherever it reliably tracks evaluatively relevant features of the context. It is not yet clear whether this is captured by our principle (4), reproduced below.

(4). Any given object is evaluated in the same way under alternative descriptions unless (a) it is not known to the agent that the descriptions (as interpreted by the agent in the context of evaluation) share an object, such that (b) it is believed to be possible that the descriptions refer to distinct objects with distinct properties, and (c) those distinct properties, if they obtained, would justify different evaluations.

In particular, it is unclear whether or not the additional information conveyed by a speaker's choice of frame affects which object is being described, or only the context in which an object is situated. For example, is the relative goodness or badness of a basketball player's shooting performance a constitutive part of the performance itself or merely of the wider context in which it occurs? The answer here depends, in turn, on how broadly or narrowly we individuate objects of description.

On a broad conception, a given object could perhaps be constituted in part by its relations to other phenomena; thus, our basketball player's performance could encompass not just the absolute proportion of shots made/missed but also whether this counts as a relatively large or small proportion (and therefore as relatively good or bad). If we were to individuate objects in this broad way, principle (4) could already accommodate the idea that agents can rationally infer and use the additional information gleaned from the speaker's choice of frame. The frames would be interpreted as having objects with (certainly or probably) different contextual relations and, therefore, as having different properties. Since relative performance properties are relevant to evaluating performance, there would be no violation of description invariance.<sup>11</sup>

On a narrower conception of objects, though, relational aspects would play no role in individuation. Instead, our basketball player's performance would be exhausted by the *absolute* proportion of shots he made or missed. Accordingly, principle (4) would render agents irrational if their evaluative judgements were

---

<sup>11</sup>This involves more fine-grained individuation of objects, then, to distinguish between choice-relevant alternatives. An analogous strategy has been put forward to deal with apparent violations of a different tenet of rational choice theory concerning the independence of irrelevant alternatives. See, for example, Broome (1993), Dreier (1996), Rulli and Worsnip (2016).

sensitive to the additional relative information conveyed by frames. This is because the frames would be interpreted as having one and the same object, which merely happens to be situated in different contexts. As argued above, I think it would still be wrong to conclude that inferring and using this contextual information is irrational, since agents' evaluations *can* reasonably depend on the additional contextual information conveyed by a speaker's choice of frame. Therefore, the characterization of the description invariance principle in (4) would no longer be fit for purpose.

Rather than adjudicating between the two metaphysical possibilities contrasted above, I propose simply to fortify the formulation of the description invariance principle, so that it is robust even on a narrow approach to object individuation. I do so by proposing one further refinement, as follows:

(5). Any given object is evaluated in the same way under alternative descriptions unless:

(a) it is not known to the agent that the descriptions (as interpreted by the agent in the context of evaluation) share an object, such that (b) it is believed to be possible that the descriptions refer to distinct objects with distinct properties, and (c) those distinct properties, if they obtained, would justify different evaluations;

or

(d) it is not known that the context of the object is the same under each of the alternative descriptions, such that (e) it is believed to be possible that there are distinct contexts with distinct properties, and (f) those distinct properties, if they obtained, would justify different evaluations.

The addition of the clauses (d)–(f) recognizes the potential relevance of contextual inferences.

It could be that this formulation of the principle will still need some further refinements to function as a foundational tenet of rational choice theory.<sup>12</sup> Nevertheless, it achieves the objective of aligning rational choice theory with recent empirical developments in the framing literature. In the next section, I briefly summarize what this means for our understanding of framing effects in general.

#### 4. Framing Effects

While myriad studies in the framing literature reveal shifts in judgements in response to linguistically distinct stimuli, it is unclear exactly what we should infer from these. I have argued here that researchers have been operating with an insufficiently refined principle of description invariance. As a result, they

<sup>12</sup>For example, one might worry that the problem of logical omniscience will arise, whereby agents could violate the principle without making a rational error, just because they fail to infer all of the implications of their beliefs about the object.

have tended to pay too little attention to how stimuli are interpreted by experimental participants and, thus, to how the objects of description and relevant contextual features are represented. As we have seen, though, this can make all the difference when it comes to the rationality or irrationality of frame-sensitive responses. Agents might be perfectly justified in making different judgements under different frames, provided that they have good reason to represent the world as (certainly, probably or possibly) different under each frame. That can happen when an object of description is represented as having different properties; and it can happen when relevant features of the context are taken to differ. The recommendation for future experimental work, then, is to test and eliminate these possibilities. This is a prerequisite for making plausible claims about violations of the principle of description invariance. Following this approach, we will be able to tell whether framing effects, understood as superficial sensitivities to merely linguistic differences, really exist, or whether the data are actually revealing agents' perfectly rational responses to distinct informational contents.

Relatedly, refining the principle of description invariance helps us assess a more radical approach to framing and rationality. Bermúdez (2020) proposes a dramatic restriction of the description invariance principle, arguing that there are many contexts in which it has no normative force. This is motivated in part by a desire to realign normative theory with actual human behaviour. However, that motivation begins to fall away as the principle of description invariance is characterized more rigorously, along the lines I have proposed. Until it has been shown that the description invariance principle genuinely *does* get violated in a systematic and pervasive way, there is little call for claiming that it *should* be, at least not as a means to bring rational choice theory closer to descriptive reality.<sup>13</sup>

In sum, by refining the principle of description invariance, it is hoped that we can sidestep the choice between an overly pessimistic picture of human reasoning (as diverging systematically and pervasively from the normative ideal) and an overly permissive picture of rationality (neutering the fundamental tenet of description invariance). That clears the way to pursuing a theory of rational choice which acknowledges both the subtleties of linguistic communication and the sophistication of human agents.

## 5. Conclusion

The rational choice principle of description invariance must be handled with care when applied to human agents with imperfect knowledge. Failure to recognise this sufficiently clearly in previous research has led to rather rash claims about our irrationality, on the basis of apparent framing effects. To remedy this situation, I have argued for a series of refinements to the characterization of that principle,

---

<sup>13</sup>Of course, this is not Bermúdez's *only* motivation for delimiting the scope of the principle of description invariance; and other aspects of his overall case need to be dealt with via separate lines of argument. See Fisher (2022a, 2022b) for further discussion.

which are designed to accommodate justifiable forms of frame-sensitivity. The upshot is a recalibration of our theorizing about human judgement and decision-making, enhancing the plausibility of our normative theory and the interpretation of recent empirical findings.

**Acknowledgements.** I thank UKRI for research support (grant reference MR/V025600/1). I am also grateful to the editor and two reviewers for this journal for their extremely helpful comments, which significantly improved the paper.

**Competing Interests.** None.

## References

- Bermúdez J.L.** 2020. *Frame It Again: New Tools for Rational Decision-making*. Cambridge: Cambridge University Press.
- Broome J.** 1993. Can a Humean be moderate? In *Value, Welfare and Morality*, ed. R.G. Frey and C. Morris, 51–73. Cambridge: Cambridge University Press.
- Chick C.F., V.F. Reyna and J.C. Corbin** 2016. Framing effects are robust to linguistic disambiguation: a critical test of contemporary theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **42**, 238–256. doi: [10.1037/xl0000158](https://doi.org/10.1037/xl0000158).
- Dreier J.** 1996. Rational preference: decision theory as a theory of practical rationality. *Theory and Decision* **40**, 249–276. doi: [10.1007/BF00134210](https://doi.org/10.1007/BF00134210).
- Fisher S.A.** 2022a. Frames, reasons, and rationality. *International Journal of Philosophical Studies* **30**, 162–173. doi: [10.1080/09672559.2022.2057685](https://doi.org/10.1080/09672559.2022.2057685).
- Fisher S.A.** 2022b. Defining preferences over framed outcomes does not secure agents' rationality. *Behavioral and Brain Sciences* **45**, e227. doi: [10.1017/S0140525X22001029](https://doi.org/10.1017/S0140525X22001029).
- Fisher S.A. and D.R. Mandel** 2021. Risky-choice framing and rational decision-making. *Philosophy Compass* **16**, e12763. doi: [10.1111/phc3.12763](https://doi.org/10.1111/phc3.12763).
- Frisch D.** 1993. Reasons for framing effects. *Organizational Behavior and Human Decision Processes* **54**, 399–429. doi: [10.1006/obhd.1993.1017](https://doi.org/10.1006/obhd.1993.1017).
- Kahneman D.** 2000. Preface. In *Choices, Values, and Frames*, ed. D. Kahneman and A. Tversky, ix–xviii. Cambridge: Cambridge University Press.
- Kühberger A.** 1995. The framing of decisions: a new look at old problems. *Organizational Behavior and Human Decision Processes* **62**, 230–240. doi: [10.1006/obhd.1995.1046](https://doi.org/10.1006/obhd.1995.1046).
- Leong L.M., C.R.M. McKenzie, S. Sher and J. Müller-Trede** 2017. The role of inference in attribute framing effects. *Journal of Behavioral Decision Making* **30**, 1147–1156. doi: [10.1002/bdm.2030](https://doi.org/10.1002/bdm.2030).
- Levin I.P., S.L. Schneider and G.J. Gaeth** 1998. All frames are not created equal: a typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes* **76**, 149–188. doi: [10.1006/obhd.1998.2804](https://doi.org/10.1006/obhd.1998.2804).
- MacDonald R.R.** 1986. Credible conceptions and implausible probabilities. *British Journal of Mathematical and Statistical Psychology* **39**, 15–27. doi: [10.1111/j.2044-8317.1986.tb00842.x](https://doi.org/10.1111/j.2044-8317.1986.tb00842.x).
- Mandel D.R.** 2014. Do framing effects reveal irrational choice? *Journal of Experimental Psychology: General* **143**, 1185–1198. doi: [10.1037/a0034207](https://doi.org/10.1037/a0034207).
- Rulli T. and A. Worsnip** 2016. IIA, rationality, and the individuation of options. *Philosophical Studies* **173**, 205–221. doi: [10.1007/s11098-015-0481-6](https://doi.org/10.1007/s11098-015-0481-6).
- Sher S. and C.R.M. McKenzie** 2006. Information leakage from logically equivalent frames. *Cognition* **101**, 467–494. doi: [10.1016/j.cognition.2005.11.001](https://doi.org/10.1016/j.cognition.2005.11.001).
- Sher S. and C.R.M. McKenzie** 2008. Framing effects and rationality. In *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*, eds N. Chater and M. Oaksford, 79–96. Oxford: Oxford University Press.
- Sher S. and C.R.M. McKenzie** 2011. Levels of information: a framing hierarchy. In *Perspectives on Framing*, ed. G. Keren, 35–63. Abingdon: Psychology Press.

- Simmons J. and L. Nelson** 2013. “Exactly”: The most famous framing effect is robust to precise wording. <http://datacolada.org/11>.
- Stanovich K.E. and R.F. West** 1998. Individual differences in framing and conjunction effects. *Thinking & Reasoning* 4, 289–317. doi: [10.1080/135467898394094](https://doi.org/10.1080/135467898394094).
- Tversky A. and D. Kahneman** 1986. Rational choice and the framing of decisions. *Journal of Business* 59, S251–S278.

**Sarah A. Fisher** is a Research Fellow in Philosophy at UCL. Her current scholarship focuses on linguistic framing and contextual effects on meaning, especially in online speech environments. URL: <https://sites.google.com/view/sarahafisher/home>.