

SYMPOSIUM ARTICLE

The Reflection Principle and the Ex-Ante Pareto Principle in Anna Mahtani's *Objects of Credence*

Luc Bovens 

Department of Philosophy, University of North Carolina at Chapel Hill, NC, USA
Email: lbovens@unc.edu

(Received 31 August 2023; accepted 16 October 2023; first published online 26 March 2024)

Abstract

First, Mahtani argues that both in the game The Mug and in the Sleeping Beauty we should not defer to a trusted person under a particular designation if they do not self-identify under this designation. This invites a more complex Reflection Principle. I respond that there are more parsimonious ways to avoid the challenges posed to the Reflection Principle. Second, Mahtani argues that preferences create a hyperintensional context, which poses a challenge to the Ex-Ante Pareto Principle that can be averted by supervaluation. I respond that such an appeal to supervaluation would block randomization as a fair allocation device.

Keywords: Rationality; Reflection Principle; Welfare Economics; Sleeping Beauty; Ex-Ante Pareto

My discussion will focus on Chapters 5 and 6 of Anna Mahtani's *Objects of Credence* (2024). In these chapters, Mahtani provides an innovative discussion of how issues in philosophy of language are relevant to paradoxes of rationality and to welfare economics. I have learned a lot from reading the manuscript and I offer my comments below in the spirit of constructive dialogue.

1. Sleeping Beauty, the Mug and the Reflection Principle

Mahtani introduces refinements to the *Reflection Principle*, that is, the principle that we should defer to our future self when assigning credences. There are the standard counter examples in which my future self is less trustworthy than my present self. So, (i) we should only defer to a *trusted* (or, in Mahtani's words, a *respected*) future self. But if trust is what matters, then we should not only defer to a trusted future self, but (ii) also to any person who is more trusted than one's present self. And we do not just defer to persons, but (iii) to persons under a designation. On literary matters, I might defer to a person under the designation of George Orwell, but not under the designation of Eric Blair, because I do not know that George Orwell and Eric Blair are one and the same person. This leads Mahtani to the following modification of the *Reflection Principle*:

© The Author(s), 2024. Published by Cambridge University Press.

Clarified Generalized Reflection Principle [CGRP]: If a rational agent-at-a-time A respects an agent at-a-time A* (so designated), then A defers to A* (so designated).

But, Mahtani continues, we need more modifications. Mahtani sides with thirders in the Sleeping Beauty (SB) Problem. On Sunday night, SB assigns credence $1/2$ to Heads. But then on Monday upon awakening, she assigns credence $1/3$ to Heads. She knows all of this on Sunday and so, since SB on Monday is a trusted self, who knows exactly what she knew on Sunday plus the additional information that there is an awakening, SB should update her credence on Sunday to $1/3$ according to the CGRP. But this is counterintuitive and so, Mahtani concludes, there is still something that is not right with the CGRP.

To find out what's wrong, Mahtani constructs a card game which she calls 'The Mug'. There are three players (Alice, Bob and Carol), and a Dealer. The Dealer randomly and secretly divides the roles of 'the mug', 'the lucky player' and 'the other player' between the players. She deals a joker to the mug, an ace to the lucky player, and flips a fair coin so that the other player gets a joker if Heads, and an ace if Tails. Now consider the proposition:

2A. Two of the cards dealt were aces.

Before the players have turned the card that is dealt to them, their credence in 2A is $1/2$. By Bayesian reasoning, once they have turned their card, they will update as follows. If they see an ace, then their credence in 2A will go up to $2/3$, while if they see a joker, their credence will go down to $1/3$. (See Chapter 5, footnote 4.)

Alice might reason as follows. Before having turned her card, her credence in 2A is $1/2$. The mug (at the time that she has turned her card) is a trusted person for Alice, since the mug has more information than she does at this point in the game. So, following the CGRP, Alice should defer to the mug, so designated, and adjust her credence to $1/3$. But clearly, Alice should not defer to the mug. So, what is wrong with the CGRP?

Mahtani adapts the CGRP as follows. The trusted person A* must be an 'appropriate designator' before Alice should defer to her. What this means is that A* identifies herself as A*, or at least, that, if she were to identify as A*, then she would continue to have the same credence in the proposition under consideration. Mahtani calls this the *Improved Generalized Reflection Principle (IGRP)*.

How does the IGRP yield the right result in the card game The Mug? Consider the mug whom Alice considers a trusted person. The mug does not know that she is the mug – she might be the other player who was dealt a joker. If she were to know that she is the mug, then her credence in 2A would shift to $1/2$, since whether 2A is true would just depend on the Dealer's coin flip. And so, Alice should not defer to the mug, so designated.

And this is also the reason that Sleeping Beauty should not defer to her awakening future Monday self. This future Monday self may be a trusted self for SB. But this future Monday self does not know that she is an awakening Monday self. For all she knows, she might be an awakening Tuesday self. And if she were told that it is Monday, then she would just shift her credence in Heads to $1/2$, since the coin is

a fair coin. And so, SB should not defer to her awakening Monday self, so designated.

This is an interesting strategy to show at the same time why we should not be deferring to a trusted future person in The Mug and how thirders in the Sleeping Beauty problem can dismiss the objection that credence assignments by SB to Heads of $1/2$ on Sunday and $1/3$ on Monday are inconsistent. But I wonder whether there may not be more parsimonious ways to deal with The Mug and the Sleeping Beauty that do not require us to adopt the *IGRP*.

In The Mug, there are four future selves which Alice should consider. She might be the mug, the lucky player, the other player who is dealt a joker, or the other player who is dealt an ace. In section 5.2.1 Mahtani writes, in response to a *Reflection Principle* that appeals to a future self, that ‘in cases where a rational agent is not certain what her credence in P will be . . . her credence in P at t_0 ought to equal her expectation of her future credence in P.’ Now in The Mug, she assigns $1/3$ credence to being the mug, $1/3$ to being the lucky player, $1/6$ to being the other player who is dealt a joker and $1/6$ to being the other player who is dealt an ace. If she turns a joker, then her credence in 2A is $1/3$ and if she turns an ace, it is $2/3$. Hence, her expected credence is $(1/3)(1/3) + (1/3)(2/3) + (1/6)(1/3) + (1/6)(2/3) = 1/2$. And so, a *Reflection Principle* that appeals to expected credences of trusted future selves in case of uncertainty does not force us to shift away from our prior credence in 2A of $1/2$.

But in Mahtani’s *CGRP* we consider not just trusted future selves, but any trusted agent-at-a-time. I take it that Mahtani restricts the *CGRP* to trusted agents-at-times whom a rational agent knows to be actual. Alice knows that there is one actual mug and one actual lucky player in the card game she is currently playing. She does not know that there is the other person who will see a joker, nor does she know that there is the other person who will see an ace. These are just possible agents-at-times for Alice. So, Alice restricts her attention to the mug and the lucky player. Deferring to the mug, Alice would lower her credence in 2A to $1/3$, while deferring to the lucky player, Alice would raise her credence in 2A to $2/3$. But since Alice should do neither, Mahtani argues, there is a problem with the *CGRP*.

That is true. But could we not make a similar move as before? The *actual* trusted agents-at-a-time whom Alice is considering are one mug and one lucky player. In the same way that we modified the *Reflection Principle* with an expectation in case of multiple trusted future selves, we could modify the *CGRP* as follows:

Clarified Generalised Reflection Principle[#] [*CGRP*[#]]: If a rational agent-at-a-time A respects a single agent at-a-time A^* (so designated), then A defers to A^* (so designated), and if A respects multiple agents-at-times (so designated), then A’s credence equals the mean of these agents-at-times’ credences.

Hence, Alice should set her credence at $(1/2)(1/3) + (1/2)(2/3) = 1/2$, that is, her credence should remain unchanged.

In the Sleeping Beauty, there is SB’s Monday future self and Tuesday future self and they each assign credence $1/3$ to Heads. Now SB’s Tuesday future self is suffering from amnesia, and this is a standard argument not to consider her as a trusted future self. But what about her Monday future self? One might argue that

prior to being put to sleep, SB knows that it is a particular day in the sequence of relevant days, viz. that it is day 0 in the sequence. Upon awakening, she does not know whether it is day 1 or day 2 in the sequence. That is, she no longer knows that it is a particular day in the sequence of relevant days. And given this loss of knowledge, SB's Monday future self, so designated, is not a trusted person. And so, the CGRP does not force her to lower her credence to 1/3.

Granted, the strength of Mahtani's appeal to 'an appropriate designator' is that it is a unified account that covers both The Mug and the Sleeping Beauty. But I am not sure that The Mug and the Sleeping Beauty *force* us to move from the CGRP to the IGRP since each case can be solved in a more parsimonious way.

2. Welfare economics, hyperintensionality and Ex-Ante Pareto

Suppose that we have a choice between two prospects. On prospect 1 (Table 1), a fair coin is flipped. If Heads, Alice receives an outcome with utility 1 and Belinda an outcome with utility 0. If Tails, Belinda receives an outcome with utility 1 and Alice an outcome with utility 0. On prospect 2 (Table 2), both Alice and Belinda receive utility 0.5 minus a small differential of ε . This choice illustrates the conflict between Ex-Ante Pareto and Egalitarianism. An Ex-Ante Paretian favours prospect 1 since both Alice and Belinda have expected utility 0.5 on prospect 1 which is greater than expected utility ($0.5 - \varepsilon$) on prospect 2. The Egalitarian favours prospect 2, since no matter what event materializes in prospect 1, the more egalitarian character of the distribution in prospect 2 offsets the small drop in average utility.

Table 1. Prospect 1

	Heads	Tails
Alice	1	0
Belinda	0	1

Table 2. Prospect 2

Alice	$0.5 - \varepsilon$
Belinda	$0.5 - \varepsilon$

Now consider prospect 3 (Table 3). On prospect 3, Alice receives utility 1 and Belinda utility 0. Ex-Ante Pareto has nothing to say about how this prospect compares with prospect 2. So, an Ex-Ante Paretian has no countervailing argument against the Egalitarian ranking of prospect 2 over prospect 3.

Table 3. Prospect 3

Alice	1
Belinda	0

Preferences over outcomes, according to Mahtani, create a hyperintensional context – that is, I may prefer an outcome that Alice is happy but not that Ms Smith is happy, even though, unbeknownst to me, Alice is Ms Smith. Notice what happens when hyperintensionality enters the choice between prospect 2 and 3. We tell the decision maker that Alice and Belinda are Ms Smith and Ms Jones, but we don't know who is who – both options are equiprobable. If we frame the choice in terms of prospects for Ms. Smith and Ms. Jones, prospect 2 is represented as prospect 4 (Table 4) and prospect 3 is represented as prospect 5 (Table 5). Note how Ex-Ante Pareto kicks in again. By bringing in some simple uncertainty about names into prospect 3, the Ex-Ante Paretian can argue that prospect 5 is preferable to prospect 4, since the expected utility of both Ms Smith and Ms Jones in prospect 5 are greater than in prospect 4.

Table 4. Prospect 4

	Heads
Ms. Smith	$0.5-\epsilon$
Ms. Jones	$0.5-\epsilon$

Table 5. Prospect 5

	Ms Smith = Alice Ms Jones = Belinda	Ms Jones = Alice Ms Smith = Belinda
Ms Smith	1	0
Ms Jones	0	1

Certainly, this cannot be right. This is Ex-Ante Pareto on steroids. Adding uncertainty about names of the people affected, Mahtani argues, cannot possibly make a moral difference to the choice of prospects.

To force Ex-Ante Pareto back into its proper place, Mahtani appeals to supervaluationism. A prospect is Ex-Ante Pareto superior *simpliciter* to another prospect if and only if it is Ex-Ante Pareto superior ‘relative to every set of designators’. So, the prospect which can be represented either as project 3 or as prospect 5 is not Ex-Ante Pareto superior *simpliciter* to the prospect that can be represented either as prospect 2 or as prospect 4, since it is not Ex-Ante Pareto superior relative to the set of designators <Alice, Belinda>.

I agree that supervaluationism provides a solution to put Ex-Ante Pareto in its proper place in the prospects above. But it may be too strong of a constraint. How so? Let us look at a case in which Ex-Ante Pareto has strong appeal on grounds of randomization as a fair allocation device. Suppose that we have a single dose of a medicine and two patients Alice and Belinda. If we split the dose, then the medicine is minimally effective – it offers a very small benefit ϵ . If we allocate the full dose, then it is maximally effective – it offers utility 1. In prospect 6 (Table 6) we flip a fair coin and allocate the medicine accordingly. In prospect 7 (Table 7) we split the dose.

Table 6. Prospect 6

	Heads	Tails
Alice	1	0
Belinda	0	1

Table 7. Prospect 7

Alice	ε
Belinda	ε

Both Alice and Belinda prefer prospect 6 to prospect 7 because it offers them greater expected utility. The decision maker ranks prospect 6 over prospect 7 on grounds of Ex-Ante Pareto. Egalitarianism has little appeal here because the benefit of a split dose is so minimal.

But now look what happens when we bring in hyperintensionality and supervaluation. Let us name the person who will get the medication ‘Ms. Lucky’ and the person who will not get the medication ‘Ms. Unlucky’. Then we can rewrite prospect 6 as prospect 8 (Table 8). Under supervaluation, the prospect that can be represented either as prospect 6 or as prospect 8 is Ex-Ante Pareto superior *simpliciter* to prospect 7 only if it is Ex-Ante Pareto superior relative to both sets of designators <Alice, Belinda> and <Ms. Lucky, Ms. Unlucky>. But it is not Ex-Ante Pareto superior relative to <Ms. Lucky, Ms. Unlucky> since Ms. Unlucky’s expected utility in prospect 8 is 0 and hence lower than ε in prospect 7.

Table 8. Prospect 8

Ms Lucky	1
Ms Unlucky	0

So, once we introduce the proper names ‘Ms. Lucky’ and ‘Ms. Unlucky’ as designators, the prospect that can be represented either as prospect 6 or as prospect 8 is no longer Ex-Ante Pareto superior *simpliciter* to prospect 7. But simply adding a pair of new designators should not take away the moral weight of the Ex-Ante Pareto principle in ranking these prospects. Hence, supervaluation is too strong a requirement. It takes the winds out of the sails of Ex-Ante Pareto as a principle that supports randomization as a fair allocation device.

Acknowledgements. I am grateful to the participants at the workshop on Anna Mahtani’s book manuscript on 22 May 2023 in the Institute of Philosophy in London. I have benefitted from subsequent discussions with Anna Mahtani and Richard Bradley.

Reference

Mahtani A. 2024. *Objects of Credence*. Oxford: Oxford University Press.

Luc Bovens is a Professor in the Department of Philosophy and core faculty in the Philosophy, Politics, and Economics Program in the University of North Carolina in Chapel Hill. He recently published *Coping: A Philosophical Guide* (OpenBook, 2021). His research interests are in Philosophy, Politics, and Economics and in Moral Psychology.

Cite this article: Bovens L (2024). The Reflection Principle and the Ex-Ante Pareto Principle in Anna Mahtani's *Objects of Credence*. *Economics and Philosophy* **40**, 478–484. <https://doi.org/10.1017/S0266267123000421>