Model behaviour: solving Mahtani's puzzles without the tenet

Harry Long

Abstract

Anna Mahtani has presented a family of puzzles in which different, comparably plausible applications of Bayesian epistemology or orthodox decision theory to the same case yield inconsistent results. Mahtani explains them by appeal to a tenet—that credence-attributions, like belief-attributions, are opaque—which she also shows to be hard to reconcile with the credence framework. Gustafsson and Kowalczyk explain them instead by appeal to trans-world identification, and leave the selection of an identification method to context. I argue for a third diagnosis. The puzzles are problems of model-selection, and the selection is governed by which model best represents the reasoner's epistemic situation rather than by what is salient to her. Two claims support this. First, puzzles of the same kind can be generated where there is no identity uncertainty at all, and hence where trans-world identification has no work to do; I give such a puzzle, generated by the reference class problem, and argue that, on plausible assumptions, it shares a root with Mahtani's own. Second, contextual selection faces a dilemma that the modelling approach avoids. I close with two heuristics constraining which well-defined problem a situation determines.

1 Introduction

Recently, Anna Mahtani has presented a family of ingenious puzzles in which different, comparably plausible applications of Bayesian epistemology or orthodox (Bayesian) decision theory to the same case yield inconsistent results (Mahtani 2017), (Mahtani 2021). Call these *Mahtani puzzles*.[1] In *The Objects of Credence*, Mahtani explains these puzzles by appeal to a claim she calls

The tenet: that just as belief-attributions are *opaque*—in that they do not necessarily retain their truth-values under substitution of designators with the same extensions and intensions—so, too, are credence-attributions (Mahtani 2024).[2]

[1] I will characterise the kind of puzzle more precisely later.
[2] Opacity, in the terms of possible worlds semantics, is *hyperintensionality*.

Mahtani gives three independent arguments for the tenet. But she also shows that it is presently unclear how the credence framework should be interpreted so as to accommodate it. We are thus left in a dilemma: either the Bayesian modelling frameworks are jeopardised by the puzzles, or we embrace a tenet whose compatibility with those frameworks is in question.

A recent alternative avoids the dilemma. Gustafsson and Kowalczyk (2026) argue that a person's prospect specifies what happens to that person across the possible states of the world, so that computing a prospect presupposes a method of identifying individuals across states. Two such methods are available in Mahtani's central case, and they yield different verdicts; the appearance of contradiction arises because the case singles neither out. Prospects are therefore relative to identification methods rather than to designators, and no revision of the credence framework is needed. The missing information, they propose, is supplied by context.

I think this diagnosis is on the right lines, but stops one step short. The question it leaves open is how a situation determines the well-defined problem to be solved in it. This is not a by-product of the puzzle; it is the puzzle.

Nor is the question peculiar to the *ex ante* Pareto principle or to Bayesian epistemology. Anyone who applies a formal model to a real situation—in decision theory, epistemic logic, game theory, or the sciences—must first get from the real-world, inchoate situation to a well-defined problem. But the situations underdetermine the problems. A decision table has states, acts and outcomes. Nothing in the world announces *which* states, which acts, and under which descriptions. Call this the *specification challenge*; §7 takes it up. The puzzles Mahtani formulates repay attention because they are an unusually tractable instance of it. The phenomenon is not confined to formal frameworks. Reasoning from examples in philosophy exhibits reversals of the same shape, since—as I defend elsewhere—a judgement elicited by a case described at one level of detail may be reversed when the same case is described at another.

Here I make three claims.

First, puzzles of the relevant kind can be generated where no individual's identity is in question, and hence where trans-world identification is idle. I give such a puzzle, generated by the reference class problem, in §4, and I argue in §4.3 that it shares a root with Mahtani's own. The puzzle shows that the phenomenon is wider than the treatment of identity across states.

Second, leaving the selection of a representation to context carries a cost its proponents have identified but, I think, underestimated. On their own account, one must either accept expectational utilitarianism or accept that facts about permissibility shift with what happens to be salient to the reasoner, with no change in her information. The metaepistemological view I will call Epistemology as Modelling

(EM)—whose components are found in (Greco 2023), (Titelbaum 2012), (Weisberg 2013), and (T. Williamson 2024)—avoids the disjunction. On EM the representation is fixed by which model best fits the reasoner's epistemic situation, and the best model *represents* her ignorance of who is who, rather than resolving it by fiat or salience.

Third, the specification challenge can be attenuated with the use of heuristics: I here defend two. The first requires that a designator occurring in the outcomes of a decision table pick out the same object at every event of that table. The second requires that uncertainty about a structural fact be internalised within a single model (rather than, say, ignored, or distributed across rival models whose verdicts are then averaged). Each heuristic is a constraint on which candidate specifications are admissible, and each is defended on grounds internal to the practice of modelling—what a prospect is, and what marginalisation is defined for—rather than imported from a theory of content.

The paper is structured as follows. §2 presents the central puzzle. §3 sets out the two existing diagnoses and what they have in common. §4 gives a Mahtani puzzle without identity uncertainty, and argues that it shares a root with Mahtani's original. §5 states EM. §6 applies it to both cases and argues against contextual selection. §7 addresses the remaining problem, and defends the two heuristics.

2 An example

Imagine that:

Surgery: You are a doctor expecting Ann and Belle—two patients with the same medical condition, d—to arrive at your clinic within the hour. You have no idea who will arrive first, and (applying the Principle of Indifference) think it equally likely that each does. You have a single dose of medication. You know little else about the patients. You can either give a full dose to whoever arrives first (the 'first come first served', or 'FCFS' policy) or divide the dose between the two (the 'split' policy). A full dose is much more effective than a half dose; so much so that it is better to have a one in two chance of a full dose, than a half dose for sure (Mahtani 2017).

Thus, you write up Table 1. Since the cell-values contain assignments of utility for

Table 1 *ex ante* Pareto superiority of FCFS, using {'Ann', 'Belle'}

Action	e_1: Ann arrives first	e_2: Belle arrives first	Prospects
FCFS	Ann: 10 Belle: 0	Ann: 0 Belle: 10	Ann: 5 Belle: 5
Split	Ann: 4 Belle: 4	Ann: 4 Belle: 4	Ann: 4 Belle: 4

each affected individual, one cannot apply orthodox decision theory to output the expected utility of each action. One can nevertheless proceed with a broadly Bayesian decision-theoretic model by calculating the prospect of each action for each individual. The *prospect* of an action for an individual is the sum of the utility of that action for that individual at each state, weighted by the reasoner's credence that each state obtains.[3] The prospect of FCFS for Ann is thus $(0.5)(10) + (0.5)(0) = 5$. An action is *ex ante* Pareto superior to another if its prospect is at least as great for everyone as, and greater for someone than, the prospect of the other. The *ex ante* Pareto principle is just that if an action is *ex ante* Pareto superior to another, then you should not do the latter. The intuitive idea is that it's impermissible to do something if doing something else is better for someone and no worse for anyone. Table 1 yields that FCFS is *ex ante* Pareto superior to splitting.

Now suppose that since arrival time matters so much, you begin to think of the patients in terms of their arrival time: coining the names 'Firsty' and 'Secondy' for the people who, respectively, arrive first and second. This is not unreasonable: *ex hypothesi*, you know little else about them. You then write out Table 2. [4] FCFS is

Table 2 No *ex ante* Pareto superiority, using {'Firsty', 'Secondy'}

Action	e_1: Ann arrives first	e_2: Belle arrives first	Prospects
FCFS	Firsty: 10 Secondy: 0	Firsty: 10 Secondy: 0	Firsty: 10 Secondy: 0
Split	Firsty: 4 Secondy: 4	Firsty: 4 Secondy: 4	Firsty: 4 Secondy: 4

not *ex ante* Pareto superior to Split: it is better for Firsty and worse for Secondy. Contradiction.[5]

Say that a set of designators is *admissible* just if there is a bijective denotation function from that set to the set of individuals affected by one's actions. The reasoner in Surgery has, and knows she has, the admissible designator sets 'Ann', 'Belle' and 'Firsty', 'Secondy', and cannot match up the referents of the one with the referents of the other. There is a partition of possibilities: (Ann is Firsty and Belle is Secondy), (Ann is Secondy and Belle is Firsty). Call these the *identity-specifications*.[6]

[3] I follow Mahtani in using 'prospect' for the expectation of welfare. Gustafsson and Kowalczyk reserve the term for the portfolio of outcomes a person faces across the states, and note that Mahtani's usage is non-standard (Gustafsson and Kowalczyk 2026). Nothing here turns on this: what matters is that either notion is defined for an individual, and so presupposes that some one individual's fortunes are being aggregated.

[4] We may even suppose that your receptionist begins to give you a body of information about Secondy: that she is (say) a divorced schoolteacher, 36 years old, has two children, lives out of town, *etc*, and that this encourages you to use the designators 'Firsty', 'Secondy' instead.

[5] I take 'Firsty' and 'Secondy' throughout as abbreviated descriptions rather than as names whose reference is fixed by description; nothing in the puzzle survives their being read as rigid, since a rigid 'Firsty' would denote one patient at both events and Table 2 could not be written.

[6] Gustafsson and Kowalczyk (2026) have rightly noted that the probabilities in the case should be read as subjective credences rather than objective chances, and that the reasoner assigns non-extreme credences to the identity-specifications.

2.1 The structure of the puzzles

We have a Bayesian problem in which incomplete information creates multiple seemingly legitimate representations, which in turn yield incompatible recommendations. Say that a *Mahtani puzzle* is any Bayesian decision-problem such that the reasoner:

1. knows all relevant descriptive property facts about the individuals affected,
2. knows all the conditional probabilities of propositions concerning those individuals that are supplied by her evidence; but:
3. lacks information of some structural kind—about identity facts, causal structure, or the joint distribution—and exactly because of this:
4. has multiple comparably plausible representations of the problem, that:
5. yield incompatible recommendations.

Two pre-emptive remarks. First, condition (3) requires that the reasoner lacks certain information. In particular, it requires that the information she lacks be such that having it would remove the multiplicity in (3). Knowledge of the identity-specifications would do this in *Surgery*. The two admissible designator sets would become inter-substitutable, and Tables 1 and 2 would collapse into one.

Second, conditions (3) and (4) requires that the reasoner's evidence not discriminate between the representations. It does not require that those representations are equally correct. In *Surgery*, neither admissible designator set is incorrect; the sets describe the same two people, and it is ignorance of who is who that leaves the reasoner without a basis for choosing. In the case of §4, by contrast, we will see that one of the two candidate partitions may be determinately the wrong one to use (since conditioning on a collider is a mistake whatever the reasoner believes). A representation may be determinately incorrect and still be one that her evidence does not exclude.

3 Two Diagnoses

3.1 The tenet

Mahtani's diagnosis appeals to the tenet. Her motivating cases are Fregean. An agent may be sure that George Orwell is a genius while doubting that Eric Blair is one, though Orwell is Blair. On the standard treatment of propositions as sets of worlds, and given that the credence function is a function, such an agent cannot be represented without contradiction. One implication Mahtani draws is that agents do not have credences about objects directly; instead, agents have such credences only about objects under designators (Mahtani 2024, p.51).

Applied to *Surgery*, the tenet entails that the reasoner cannot assign a credence to the proposition that a particular individual arrives first independently of how that individual is designated. Since the prospects of the actions are defined in terms of such credences, and the facts about *ex ante* Pareto superiority are defined in terms of the prospects, those facts inherit the designator-relativity. The contradiction is

generated by the reasoner's use of two different admissible designator sets.

Adopting the tenet engenders a stubborn dilemma. If the states in Ω are metaphysically possible worlds, it becomes unclear what the objects of credence are, since credence-attributions can no longer be understood as picking out the set of worlds at which some n-tuple stands in some relation. But if the states are *not* possible worlds—sets of sentence-worlds, say, or sentence-context-worlds—then as Mahtani herself acknowledges, familiar problems from ambiguity and the need for an artificial language make any defensible account elusive (Mahtani 2024). We are thus in a bind. Either the Bayesian modelling frameworks are jeopardised by the puzzles, or we embrace a tenet whose compatibility with those frameworks is in question.

3.2 Trans-world identification

Gustafsson and Kowalczyk (2026) offer a diagnosis that incurs neither cost. Their starting point is that a person's prospect specifies what happens to that person *across* the states, in the manner of a probability-weighted portfolio. Determining a prospect therefore presupposes *some way* of identifying individuals across states.

In the version of *Surgery* they consider, the reasoner has forgotten which of her patients is surnamed 'Smith' and which 'Jones', while knowing that it was Smith who arrived first. Two identifications are then available. One identification proceeds by first name and medical condition, and one by surname and arrival time. Relative to the first, each patient faces a prospect of a one-in-two chance of a full dose, and FCFS comes out as *ex ante* Pareto superior to splitting. Relative to the second, one patient is certain to receive a full dose and the other certain to receive nothing, and FCFS does not come out as *ex ante* Pareto superior to splitting. The verdict thus turns on the identification method employed. The case as described underdetermines the method. Prospects need not be relativised to designators, as Mahtani proposes; they are *already* relative to identification methods.

Two important commitments undergird Gustafsson and Kowalczyk's approach. First, the probabilities are (subjective) credences rather than objective chances, and the credences attach to epistemic possibilities—ways the world might be, for all one knows—rather than to metaphysical possibilities. This is because only epistemic possibilities can differ with respect to identity facts.[7]

The second commitment is an argument that trans-world identities cannot be read off the names used in describing the case. Call it the *transitivity argument*. The case requires two epistemic possibilities: one at which Ann is Smith, and one at which Ann and Smith are distinct individuals. Suppose we identify across the two by name. The individual at the first possibility is Ann, and so is identical with the individual called 'Ann' at the second; that same individual is Smith, and so is identical with the individual called 'Smith' at the second. But those two individuals at the second

[7] The distinction is discussed in (Chalmers 2011). Gustafsson and Kowalczyk observe that without it the puzzle appears unable to arise at all.

possibility are distinct, and identity is transitive. Identification by name is therefore unavailable. Nor, given the symmetry of the case, is there ground for identifying the individual at the first possibility with Ann rather than with Smith, or the reverse. Gustafsson and Kowalczyk conclude, with Chalmers, that the individuals populating one epistemic possibility do not in general populate others: properties are re-identifiable across epistemic possibilities, and objects are not.

Their answer to the question of how an identification gets singled out is that context supplies it by way of the counterpart relation. Since resemblance is context-dependent, so is the operative relation. This apparatus yields three explanations. A shift in the operative relation mid-puzzle explains why each step seemed plausible where it appeared. It explains the failure of substitution of co-referring names in claims about prospects. And it explains the intuition that additional information about a patient can change whose prospects one attends to—indeed, they note, a shift may be prompted by nothing more than a change in what is salient, as when one of the patients telephones ahead.

3.3 What the two diagnoses share

The two diagnoses disagree about where the relativity lies, and agree that it lies somewhere in the semantics or metaphysics of prospect-ascription. On Mahtani's view, the bearer of the relativity is the designator. On Gustafsson and Kowalczyk's, it is the identification method. Both then face a *selection question*: of the designator set or the identification method. Neither treats that question as belonging to the theory of prospects. Mahtani canvasses readings of the quantifier in the *ex ante* Pareto principle; Gustafsson and Kowalczyk hand the question to context.

My proposal is that the selection question is the whole of the problem, and that it is a question, ultimately, about *models* (of the far more inchoate real-world phenomena)! If I am right, then a theory of how situations determine representations is what the puzzles call for. A theory of the objects of credence is not. The proposal needs two things. It needs a reason to think that the choice of a designator set and the choice of a reference class are choices of the same kind, which I give in §4.3. And it needs an account of how such choices are made, which is the business of §§6 and 7. I take these in order, beginning with a puzzle in which identity is no part of the question.

4 A Mahtani puzzle without identity

Simpson's paradox occurs if the direction of association that exists overall is reversed or absent for every cell of some partition. There might be a negative association overall between *womanhood* and *admittance* to Berkeley and yet, also, a positive association between *womanhood* and *admittance* for every subject an applicant can possibly apply for (Bickel, Hammel, and O'Connell 1975). All that's required is that women apply disproportionately to a highly selective subject.

To take another example, let A, B, C and D be dichotomous variables with the

following values: a = survives, a' = dies, b = administered the treatment, b' = administered the placebo, c = has small kidney stones, c' = has large kidney stones, d = is under-fifty, d' = is fifty or older. Call B the 'treatment' variable, A the 'response' or 'success' variable, and C, D 'partition' variables. Consider the data captured by the following contingency tables (Tables 3, 4 and 5) which count, in their cell-values, the absolute number of individuals with the corresponding features:

Table 3 Treatment data partitioned by B only

	Treatment (b)	Placebo (b')	Total
Success (a)	5,475	25,250	30,725
Failure (a')	45,025	29,750	74,775
Total	50,500	55,000	105,500

Table 4 Data partitioned by B and C

	Treatment (b) Small (c)	Placebo (b') Small (c)	Treatment (b) Large (c')	Placebo (b') Large (c')	Total
Success (a)	5,000	250	475	25,000	30,725
Failure (a')	45,000	4,750	25	25,000	74,775
Total	50,000	5,000	500	50,000	105,500

Table 5 Data partitioned by B and D

	Treatment (b) <50 (d)	Placebo (b') <50 (d)	Treatment (b) ≥ 50 (d')	Placebo (b') ≥ 50 (d')	Total
Success (a)	25	500	5,450	24,750	30,725
Failure (a')	475	4,500	44,550	25,250	74,775
Total	500	5,000	50,000	50,000	105,500

For instance, there are 5,475 individuals who are both treated (b) and survive (a). These tables exhibit Simpson's paradox with respect to the partition variable C (inequalities (1), (2), (3)) because the data is oddly distributed in a particular way (inequalities (4), (5)). That is, there is a negative association overall between the treatment and survival, but a positive association between the treatment and survival

for every stone size. The aggregate-level negative direction of association seen in (1) is retained for a different partition D, as we see in inequalities (6), (7):

$$P(a|b) < P(a|b') \quad (1)$$
$$P(a|bc) > P(a|b'c) \quad (2)$$
$$P(a|bc') > P(a|b'c') \quad (3)$$
$$P(a|c) < P(a|c') \quad (4)$$
$$P(c|b) > P(c'|b) \quad (5)$$
$$P(a|bd) < P(a|b'd) \quad (6)$$
$$P(a|bd') < P(a|b'd') \quad (7)$$

Now,

Simpson: Suppose the actual frequencies in the study, your own credences, and the objective conditional probabilities are identical, and such that the inequalities (1) through (7) come out as true according to them (call such probabilities 'Simpson's paradoxically structured').[8] Your evidence consists of the conditional probabilities recorded in Tables 3, 4 and 5, and you know exactly which reference class each affected individual belongs to with respect to all relevant partition variables (including C and D). What you do not know is the causal structure underlying the conditional probabilistic information. Sophie is an individual with small kidney stones (c) under fifty (d). You have to either treat her (b) or not (b'). The utility associated with survival (a) is 1, and that associated with death (a') is 0.

There are two related problems here: the reference class problem, and the decision-problem. Consider the first.

Let X be a proposition attributing a particular property to some individual (such as the proposition that John Smith—a consumptive Englishman visiting Madeira—lives to 60, or the proposition that Sophie survives). The *reference class problem* obtains if the reasoner has difficulty determining what the unconditional (total) probability of X is, when all she has are many conditional probabilities of the form P(X, given A_1), P(X, given A_2), P(X, given A_3), *etc*, that differ from each other, and that may—but need not—be informed by the statistical evidence.[9] Each proposition A_i places the object described by X into a reference class; but since the A_i are consistent, they do not form a partition. Consequently, the Law of Total Probability—which equates P(X) to P(X, given A_1) + ... + P(X, given A_n) but *only if* A_1,...,A_n form a partition—cannot be applied.

Consider now the decision problem of what to do in *Simpson*. There are three

[8] The point does not depend on *identity* of the frequencies, the reasoner's credences, and the chances; I am equating them only for simplicity. Ordinarily, statistical evidence (in the form of actual frequencies) informs the (objective or subjective) conditional probabilities, but cannot generate them all on its own without bridge principles.

[9] The problem is epistemological. The inference the reasoner wants to make here is called the *direct inference*.

families of framework with which to solve it: Savage-style decision theory (Savage 1954), Jeffrey-style evidential decision theory (EDT) (Jeffrey 1965) and causal decision theory (CDT) (Pearl 2009). The events are not probabilistically independent of the actions, so Savage's framework is inapplicable. Applying causal decision theory requires knowledge of the causal structure, which the reasoner lacks. Evidential decision theory requires only the conditional probabilities she has, and is therefore the natural framework to apply.

Suppose you begin by thinking of Sophie as a patient with a certain stone size, and place her in the reference class of small-stoned patients. You generate Table 6.

Table 6 Decision with Sophie placed in the *c*-class

Action	*a*: Sophie survives	*a'*: Sophie dies	EU
Treat (*b*)	1	0	0.10
Don't treat (*b'*)	1	0	0.05

For instance, the $EU(Treat)$ is reached as follows:

$Cr(a|bc)U(a) + Cr(a'|bc)U(a')$

$= 0.10(1) + 0.90(0) = 0.10$

You should treat. But suppose you then think of Sophie as a patient with an age, and place her in the reference class of the under-fifties:

Table 7 Decision with Sophie placed in the *d*-class

Action	*a*: Sophie survives	*a'*: Sophie dies	EU
Treat (*b*)	1	0	0.05
Don't treat (*b'*)	1	0	0.10

You should not treat: contradiction. Sophie falls into multiple reference classes, and it is not clear how to back out either the total probability that she survives, or the expected utility of treating her.

4.1 Why the finest class is not available

An initially appealing response is to place Sophie in the finest relevant class—to use the cell of the variable obtained by taking the Cartesian product of all the relevant partition variables, here the class of small-stoned under-fifties. Call this the *finest relevant partition strategy*. It might seem that evidential decision theory implies it. On Jeffrey's framework the desirability of an act is fixed by the agent's credences conditional on that act, where her credence function is already conditional on her total evidence. She has no latitude about what to condition on. Were her credence function defined on the proposition that Sophie survives, given treatment, small stones and youth, that would settle the matter.

The strategy fails for two independent reasons. The first is that her evidence is the study, and the margins of a study do not determine its joint distribution. Tables 3, 4 and 5 are jointly consistent with any value whatever for the probability that an untreated small-stoned under-fifty patient survives.[10] There is thus no finest-class conditional for her to use, and she is thrown back on the broader classes that conflict. This is the conflict between narrowness and precision familiar from work on direct inference (Wallmann 2017; Thorn 2017).

The second reason is that, even were the finest class well populated, conditioning on everything one knows is not in general a way of finding out what one's act would do. As Pearl and others have shown on the basis of Cartwright's work, the finest relevant partition variable may be a mediator or a collider, and may therefore be such that it should *not* be conditioned on in either the direct inference or the decision-theoretic calculation (Cartwright 1979; Spirtes, Glymour, and Scheines 2000; Pearl 2009).[11] Which covariates may legitimately be conditioned on is fixed by the causal structure, and the reasoner does not know the causal structure, *ex hypothesi*.

One might ask whether knowing that stone size is a mediator or a collider would give an evidential decision theorist a reason to condition on less. In general it would not. Learning that an association is non-causal does not make it less evidential. News value is what evidential decision theory tracks; cases of this shape are exactly those in which EDT is thought to give the wrong recommendation, and the lesson usually drawn is that one should doubt the theory rather than revise how it is applied.

---

[10] Writing x for the number of treated patients who are small-stoned, under fifty and survive, and y for the number who are small-stoned, under fifty and die, Tables 3, 4 and 5 jointly entail that x lies in $[0, 25]$ and y in $[450, 475]$, so that $P(a|bcd)$ lies in $[0, 25/475]$, that is, in $[0, 0.053]$. They place no constraint on the corresponding untreated conditional: $P(a|b'cd)$ may take any value in $[0, 1]$. The constraint is thus asymmetric, and the sign of the difference—which is what the decision turns on—is left open.

[11] In the *mediated* structure, treatment directly promotes recovery and indirectly impedes it by causing stubborn small stones, which are so counter-conducive to recovery that the overall negative association of (1) results (Hesslow 1976). In the *collided* structure, the placebo in fact promotes recovery, and receiving the placebo and recovering each cause one to have small stones; conditioning on each value of C then yields the associations in (2) and (3) despite (1).

4.2 The puzzle is of the same kind as *Surgery*

Simpson satisfies the characterisation of §2.1. The reasoner knows all relevant descriptive property facts about the individual affected and all the conditional probabilities her evidence supplies. She lacks structural information, in this case, the causal graph; because of this she has two comparably plausible representations. These yield incompatible recommendations. Knowledge of the graph would remove the multiplicity, by telling her which variables to use.

Simpson lacks any question about who is who. There is one affected individual. Her identity is not in doubt, no designator in the tables varies its reference across the events, and there is nothing for a method of trans-world identification to do. The representational choice concerns which of two overlapping reference classes to place her in, and the states of the decision tables are just her survival and her death.

This bears on the containment argument by which Gustafsson and Kowalczyk limit the damage their puzzle does. They argue that the opaque-identity puzzle does not arise for Harsanyi's version of the *ex ante* Pareto principle, which concerns objective chances rather than credences, nor for preference-based versions; and hence that decades of work in welfare economics and ethics are less threatened than Mahtani suggests. The argument works by locating the puzzle's source in the difference between epistemic and metaphysical possibility. But it has no purchase on *Simpson*, where there are no identity facts about which epistemic and metaphysical possibilities could differ. The containment of the one puzzle is therefore a reason to attend to the other: what survives containment is the general phenomenon of which the opaque-identity puzzle is one instance.

The case can be strengthened so as to engage the very principle under containment. Add Tom, a patient with large kidney stones who is under fifty, and let the available policies be treating both patients and treating neither. Placing each patient in the class fixed by stone size, treatment is better for Sophie and better for Tom; placing each in the class fixed by age, no treatment is better for either. Treating both is therefore ex ante Pareto superior to treating neither on the first representation, and treating neither is ex ante Pareto superior to treating both on the second. The *ex ante* Pareto principle thus delivers a claim and its negation in a case containing no identity uncertainty at all.

4.3 The relation between *Surgery* and *Simpson*

Satisfying a common characterisation is weak evidence of a common root, since the characterisation of §2.1 is partly disjunctive. Something stronger is available. On certain assumptions, designator-selection and reference-class-selection are one problem rather than two.

Notice first how similar the two difficulties look when stated side by side. In *Surgery*, many designators pick out each affected individual, and the reasoner cannot see how

to back out the total probability that a given individual arrives first. In *Simpson*, Sophie falls into many reference classes, and the reasoner cannot see how to back out the total probability that she survives. In each case the reasoner has an array of conditional probabilities and no way of combining them, because the conditions do not form a partition. The following thesis makes the similarity exact.

The Partition Assignment Thesis: in using the designators of an admissible designator set to compute the prospects of the available actions, the reasoner assigns the individuals so designated to cells of a partition variable.

Suppose the thesis is true. What a prospect calculation requires is a total probability: the probability that the individual to whom the prospect is being assigned arrives first. Call it ρ. Each admissible designator set assigns the affected individuals to the cells of a partition variable, and ρ is well defined only relative to such an assignment. Relative to the assignment given by {'Ann', 'Belle'}, ρ is 0.5 for each of them, and Table 1 follows. Relative to the assignment given by {'Firsty', 'Secondy'}, ρ is 1 for one and 0 for the other, and Table 2 follows. Nothing in the reasoner's evidence selects an assignment, and the law of total probability will not combine the two values, since the propositions describing the assignments are consistent with one another and so do not partition.

That is the reference class problem in its usual form. Sophie's difficulty in *Simpson* is that the probability of her survival if treated is 0.10 relative to the assignment by stone size and 0.05 relative to the assignment by age. The reasoner's difficulty in *Surgery* is that the probability that a given patient arrives first is 0.5 relative to one assignment and 1 or 0 relative to the other. In each case a total probability is well defined only once the affected individual has been assigned to a cell, and in each case the evidence leaves the assignment open.

I do not defend the Partition Assignment Thesis at length, and the argument of this paper does not require it. Without it, §2.1 still classifies the two puzzles together, and §§6 and 7 still apply to both. What the thesis adds is an explanation of why the classification is apt. Two considerations support it.

The first is descriptivist. A central theme of that tradition is that designators are associated by speakers with collections of descriptions that determine their referents (Evans 1982). A description, or conjunction of descriptions, assigns an n-tuple to exactly one cell of some partition variable: by describing an object as possessing a collection of properties, the description assigns that object to a cell of the variable generated by taking the Cartesian product of those properties. One need not hold that designators are analysable by appeal to such descriptions, nor subscribe to the descriptivist programme wholesale, to hold that designators are typically associated with them. If they are, the thesis follows.

It is worth being explicit that this support does not smuggle the tenet back in. The

tenet is a claim about the objects of credence: that credence-attributions do not survive substitution of co-designating terms. The Partition Assignment Thesis is a claim about what a reasoner does when she picks a designator set for use in a decision table. It is silent on what her credences are about.

The second consideration is that an independent formal notion does the same work. Aloni's conceptual covers are sets of individual concepts that identify every individual exactly once at every world, and a cover therefore induces a partition of the relevant domain at each world.[12] Gustafsson and Kowalczyk's identification methods play the same role in their treatment of *Surgery*. An admissible designator set behaves, on the Partition Assignment Thesis, as a conceptual cover does; the question which cover is operative becomes just the question which designator set to use. That three literatures have converged on a notion of this shape is some evidence that the notion is tracking something common.

If the Thesis is true, then, the *Surgery* and *Simpson* have an identical root. In both, an individual belongs to cells of several partition variables. In both, nothing in the reasoner's evidence selects among them; and in both, the quantity she needs depends on which she selects. *Surgery*'s ignorance of the identity-specifications, and *Simpson*'s ignorance of the causal graph, are two ways of failing to know which assignment to make.

5 Epistemology as Modelling

I formulate EM as the conjunction of the following claims.[13]

Unavoidability: As soon as we talk of folk psychological entities—such as beliefs and instances of knowledge—or Bayesian epistemological entities—such as credences—we are *already* engaged in modelling (within a framework); and there is no way of doing the former without the latter (Greco 2023), (Titelbaum 2012).

Interpretivism: For a given context C and person S, for S to have credence n that p is for the *best Bayesian model* of S in C to be one in which she's represented as having credence n that p by that model. The best such model has maximal simplicity, strength, and fit with data (Greco 2023, p.99).[14]

Specification Interpretivism: for S to face decision-problem P in C is for the best model of her situation in C to represent her as facing P.

Model Contextualism: Provided there is no (semantic) modelling framework that can

[12] See (Aloni 2001, pp.105-8), whose account draws on Hintikka; and, for counterpart-theoretic treatments of the same phenomenon in quantified epistemic modality, Gerbrandy (2000) and Ninan (2018).

[13] The view has further components inessential to the argument. Following Titelbaum (2012) I distinguish modelling frameworks—such as the credence framework, which represents an agent's degrees of belief with a probability space—from models, which instantiate them.

[14] Greco is also an Interpretivist about belief, and when I speak of his Interpretivism in what follows, I will assume the view encompasses both beliefs and credences.

presently offer a perfect model of the domain in question, we are permitted to use different semantic models in different contexts (Greco 2023, p.70).[15]

Proponents of *EM* use the view to explain phenomena in formal epistemology that have led others to adopt *logically*—but not necessarily *explanatorily*—stronger alternate views.

Adopting *EM* as a working metaepistemology entails certain costs and benefits. Against Mahtani, *EM* is the cheaper option. The tenet requires an account of the objects of credence that she herself shows to be presently unavailable, and *EM* requires no revision of the credence framework at all. Against Gustafsson and Kowalczyk the accounting is reversed. Their solution needs no metaepistemology and mine needs this one. What the additional commitment yields, though, is generality. Their answer to the selection question is an answer about identifying individuals across states, and it is silent where no individual's identity is in question. *EM* answers the same question in a form that reaches *Simpson*, where trans-world identification has nothing to say, and the specification challenge beyond it. Whether that is worth the commitment is a comparative judgement. §§6 and 7 are meant to inform it.

6 Model-selection

6.1 *Surgery*

Consider what the reasoner in *Surgery* knows. She knows some facts about the patients under one pair of designators and other facts under the other, and she does not know how the pairs match up. A model built on either admissible designator set alone represents her as knowing which individuals the designators in that set pick out. It therefore fails on fit, whatever its merits with respect to simplicity and strength. The model that fits distributes her credence over the identity-specifications and records what happens to Ann and to Belle at each. It returns a single verdict about *ex ante* Pareto superiority.

It matters which of the two designator sets is held fixed, and fit decides that too. The reasoner knows which of her patients is Ann and which is Belle; what she does not know of either is whether she arrives first. A model whose states hold the patients fixed and vary the arrival order represents exactly this. A model whose states hold Firsty and Secondy fixed must vary something else she is uncertain between, and since arrival order is what she is uncertain about, there is nothing left for its states to vary but which patient bears which name—which she knows. It therefore represents her as ignorant of something she is not, and fails on fit. The same holds of the surnames in Gustafsson and Kowalczyk's variant.

[15]Three objections to *EM* deserve more attention than I can give them here: that it yields an odd partition between things that exist because the best model says so and things that exist anyway (Greco 2023, p.33); that it leans on theoretical virtues whose nature is itself contested, with a threatened regress if Interpretivism is applied at that level too; and that in many contexts it is implausible to attribute one precise credence function rather than a slightly different one. But none of these is knockdown, and none bears on the comparative claim advanced here.

Note what this argument does not do. It would be too easy to argue that a model generating a contradiction is unlikely to be the best model of the agent. Inconsistency is not as such a failure of simplicity, strength or fit; a reasoner with inconsistent dispositions may well be best modelled as inconsistent, and an argument of that form would dissolve puzzles that ought to be taken seriously. The failure identified here can be stated without mentioning the contradiction. A model that selects one designator set represents the reasoner as possessing information she lacks.

By *Specification Interpretivism*, the decision-problem the reasoner faces is the one that model represents her as facing. The question what she ought to do in *Surgery* accordingly has a determinate answer. There is a second reason to prefer that model, independent of the verdict it returns. The *ex ante* Pareto principle quantifies over the individuals affected, so a designator used to compute a prospect must pick out the same individual at every event of the table in which it occurs; otherwise the weighted sum defining the prospect is not a sum of utilities accruing to any one person. 'Ann' meets this condition. 'Firsty' does not, since it picks out Ann at e_1 and Belle at e_2. The quantities entered in Table 2 are thus not the prospects of any individual, and the *ex ante* Pareto principle has nothing to take as argument. Therefore, it outputs no verdict. I return to this constraint in §7.2.

6.2 Why the states may be read as metaphysical possibilities

The argument of §6.1 presupposes that Ann and Belle are available for re-identification across the events of Table 1. Two arguments deny that they are. The transitivity argument of §3.2 denies it for the case at hand. Mahtani denies it generally: 'George Orwell' and 'Eric Blair' are both rigid and in fact co-refer, so they co-refer at every metaphysically possible world at which their bearer exists, and yet a decision-maker may have positive credence that Orwell is not Blair; a state is needed at which that holds, and no metaphysically possible world will do (Mahtani 2021, p.2087-8). If either argument succeeds for *Surgery*, the model described in §6.1 is unavailable and an identification must be supplied from elsewhere.

The two arguments share a structure. Each requires that some designator shift its reference across the states. The transitivity argument requires a state at which Ann is Smith and a state at which she is not, which is possible only if 'Ann' or 'Smith' denotes differently at the two, or if the two are populated by distinct individuals bearing the same names. Mahtani's argument requires a state at which Orwell is not Blair, which is possible only if one of those names denotes differently across the states.

Neither requirement is forced on us in *Surgery*. The reasoner's uncertainty in *Surgery* is uncertainty about contingent matters: who arrives first, and which patient bears which surname. Each is settled differently at different metaphysically possible worlds, and Ann and Belle exist at all of them. Read that way, the states of Table 1 are ordinary sets of metaphysically possible worlds, the patients are re-identifiable across them as a matter of course, and no identification method needs selecting. What varies

across the states are the descriptions—'Firsty', 'the one called Smith'—which is a familiar fact about non-rigid designators and not a puzzle about identity. Gustafsson and Kowalczyk's own qualitative description of the case supports this reading, since it characterises each individual partly by what she is called.

Mahtani's argument does establish, however, that no single model will serve every purpose. A decision problem in which the reasoner is uncertain whether Orwell is Blair cannot take metaphysically possible worlds as its states. By Unavoidability it was never open to us to work outside a model at all, and by Model Contextualism a framework that serves in one context need not serve in all. The question is therefore not whether the states of decision problems are epistemic possibilities *tout court*, but whether the best model of a given reasoner in a given context represents them as such.

Recall that Gustafsson and Kowalczyk solution had two parts. First, that a prospect presupposes a method of identifying individuals across states. Second, that context supplies the method. I accept the first. The question is whether the second is needed. They face a dilemma.

Suppose the states of *Surgery* can be read as metaphysical possibilities. Then the transitivity argument has no premiss to stand on, since 'Ann' does not shift its reference. Ann and Belle are re-identifiable across the states as a matter of course, and the identification the first part requires is supplied by the case rather than by context. The counterpart-theoretic apparatus is then idle, and §6.1 settles the representational choice without it. But suppose instead that the states of *Surgery* cannot be read as metaphysical possibilities. Then the apparatus is needed, and with it comes the cost set out in §6.3.

So: either the machinery is unnecessary for these cases, or it is necessary, *and costly*. Gustafsson and Kowalczyk take the epistemic-possibility assumption over from Mahtani and acknowledge that their puzzle could not arise without it; T. Williamson (T. Williamson 2021, pp.311-4) notes a similar reaction. What is new here is the claim that the uncertainty in these cases concerns contingent matters, and hence that the assumption is not needed for them. I take the first horn. The next subsection shows what the second costs.

6.3 The cost of contextual selection

Suppose, then, that an identification must be supplied by context. Gustafsson and Kowalczyk propose that it is supplied by a counterpart relation, and since resemblance is context-dependent, so is the operative relation. The cost is one they state and, I think, underrate. If context can shift with what is salient, then the antecedent of the *ex ante* Pareto principle can hold in one context and fail in another with no change in the reasoner's information.

They observe that this need not infect permissibility. Suppose we hold, with Mahtani, that the value of a person's prospect is its expected welfare. Then notice what an

identification does and does not alter. Each identification takes the same distribution over outcomes and divides it among persons differently: one gives Ann and Belle five apiece, the other gives Firsty ten and Secondy nothing. What is divided is the same in each case, since the total expected welfare under an option is fixed by the outcomes and their probabilities, and not by who counts as the same person across states. Under FCFS that total is ten either way; under Split it is eight either way.

Two consequences follow. A criterion that ranks options by the total expected welfare returns the same verdict whichever identification is operative. And any criterion sensitive to how that total is distributed among persons must vary with the identification, since the distribution is precisely what the identifications disagree about—which catches the ex ante Pareto principle, and ex ante egalitarian and prioritarian principles besides. So if permissibility is to be context-invariant, the operative criterion must be insensitive to distribution; and the distribution-insensitive criterion is expectational utilitarianism. Gustafsson and Kowalczyk present this as a way of streamlining a version of Harsanyi's argument for utilitarianism.

Conditional or not, the resulting disjunction is uncomfortable. Either one accepts expectational utilitarianism, or one accepts that permissibility claims can turn on what happens to be salient. The first is a substantive and contested ethical commitment. A treatment of a puzzle about representation should not require it, and should not deliver Harsanyi's conclusion as a by-product. The second is a significant concession.

EM avoids the disjunction where the first horn is available. On EM the representation is fixed by which model best fits the reasoner's epistemic situation, and the salience of a patient's telephone call does not change what she knows. Where, as in Surgery, the states can be read as metaphysical possibilities, the model that fits represents her ignorance of who is who and returns an invariant verdict.

It may be objected that EM is no better off, since Specification Interpretivism makes the problem a reasoner faces relative to a context C; if a shift in salience is a shift in C, then permissibility turns on salience here too.

Everything depends on what fixes C. Two of Gustafsson and Kowalczyk's own variants separate the positions. In the first, one of the patients telephones the clinic, apologises for running late, and rings off before the reasoner can ask for her first name. The reasoner learns nothing about who is who; what has changed is only that Jones has become *vivid* to her. On their account this may shift the operative counterpart relation, and so the verdict. On EM it shifts nothing, because what a model must fit is the facts about what the reasoner knows and does not know, and the call leaves those untouched. In the second variant the receptionist produces thick dossiers on Smith and Jones—their education, projects and careers—containing no first names. Here what she knows *has* changed, so on EM the best model of her may change with it. A counterpart relation fixed by resemblance in contextually salient respects shifts in

both variants alike, and so cannot mark the difference between them. That asymmetry is what EM generates: the model moves with information, and not with salience.

I do not claim that the epistemic-possibility framing is never the right one. The claim defended here is correspondingly scoped. Take a genuine Fregean case, in which the uncertainty concerns a necessary identity: the reasoner wonders whether Hesperus is Phosphorus. There is no metaphysically possible world at which they differ, so the first horn of the dilemma is unavailable, and states finer than possible worlds are needed. There EM has no more to offer against salience-sensitivity than its rivals do; what Model Contextualism adds is only that a finer-grained model may be used in such a case without thereby committing us to the tenet as a general thesis about the objects of credence.

The claim is therefore comparative, and it concerns the cases in this family. Whether Ann arrives first, and which patient bears which surname, are contingent matters, settled one way at some metaphysically possible worlds and another way at others. For those cases the simpler framing is available, it preserves determinate verdicts, and by Interpretivism it is to be preferred. What generates the puzzle is the importing of machinery designed for Frege cases into cases that are not Frege cases, which may be an instance of what Williamson calls overfitting (§2.6, T. Williamson 2024).

6.4 *Simpson*

The treatment of *Simpson* runs along the same lines, and the substantive work is done by causal modelling rather than by EM directly. There are Bayesian frameworks that take causal uncertainty into account. Wallmann and Williamson (2017) propose that, given statistical, causal and mechanistic information, the reasoner construct a Bayesian net satisfying the Markov condition, and apply an inference algorithm to infer the maximum-entropy probability function compatible with the resulting joint distribution. That function assigns a total probability to the proposition that Sophie survives, and with a suitable utility function it can be used to partition the states before applying causal decision theory.

I do not claim that this solves the reference class problem in general. I claim only that mature, causally sophisticated Bayesian modelling can in principle supply a single, highly plausible model in place of two that conflict, and that our puzzlement dissolves to the extent that we model the case better. Where the reasoner is uncertain that her net contains the right mechanisms and variables, the Bayesian paradigm allows her to set priors over the candidates and propagate the uncertainty through to the estimate. EM's contribution is Specification Interpretivism: adopting the best such model settles which problem she faces, and thus what she ought to do.

The structural parallel with *Surgery* runs as follows. In *Surgery*, ignorance of the identity-specifications leaves the reasoner unable to assign the affected individuals to cells of one partition variable rather than another. In *Simpson*, ignorance of the causal graph leaves her unable to assign the affected individual to one such cell rather

than another. In both cases, the response is to build the ignorance into a single model rather than to resolve it arbitrarily. The difference is that in *Simpson*, there is no identity question, and so no answer in terms of trans-world identification to be had.

7 The specification challenge

7.1 The challenge

It may seem that Mahtani puzzles are rarefied curiosities of formal epistemology. It is more plausible that they belong to a general challenge facing anyone who wants to use a formal model—of Bayesian epistemology, decision theory, epistemic logic, game theory, or the sciences—to arrive at the truth.

§1 introduced the *specification challenge*: the challenge of describing how one gets, from a situation, to the small-world problem to be solved in it, rather than to any number of candidate problems with correspondingly different solutions. Certain formal problems enjoy a one-one relation between problem and solution; determine a probability space $<\Omega, F, P>$ and there exists by definition a single posterior. But situations underdetermine such problems (Rothschild 2022). The *Surgery* case description does not on its own imply Table 1 or Table 2, and the *Simpson* case description does not on its own imply Table 6 or Table 7.

Two aspects can be distinguished. The *representation selection challenge* is that of ascertaining which aspects of the situation comprise the relevant objects, properties and relations; an astronomer may sensibly include the planets, and not air resistance, in the problem of predicting elliptical paths.[16] The *granularity challenge* is that of ascertaining the degree of detail to which the situation should be construed. Gustafsson and Kowalczyk's question about how the missing information is supplied is an instance of the first. What follows here is a partial answer for the decision-theoretic case: two heuristics which, other things equal, any acceptable problem-specification will satisfy.

7.2 Heuristic 1: Constancy across events

Suppose that:

Two Envelope:You must choose between two envelopes each containing a cheque, one for twice the amount of the other, where the range of possible amounts is bounded and discrete. You receive only the sum in the envelope you pick, and have no idea which envelope is which. You pick one at random. Should you switch?

Let 'M' stand for the amount of money actually in whichever envelope you picked, and 'N' for the amount of money in the other envelope:

[16]Meeting it may require analogues of the composition, ordering, weighting, supplementation and deformation operations described in Goodman (1978); see also (Chapter 3, Weisberg 2013).

Table 8 Two-envelope: Switch recommended

	e_1: N = 2M	e_2: N = 0.5M	Expected Utility
Stick	M	M	M
Switch	2M	0.5M	1.25M

Since you have no idea which envelope is which, your credence that event e_1 obtains = your credence that e_2 obtains = 0.5. Maximising expected utility—as orthodox decision theory recommends—you should switch. But our situation, though, appears to be equally well represented by Table 9. If we maximise expected utility, we should

Table 9 Two-envelope: Stick recommended

	e_1: N = 2M	e_2: N = 0.5M	Expected Utility
Stick	0.5N	2N	1.25N
Switch	N	N	N

Stick, and hence not Switch: contradiction.

Mahtani notes that the puzzle is generated by variation in the expected value of the designator 'M' occurring in the outcomes. The value of 'M' in the first column of Table 8 is *half* the value of 'N', and in the second column, *twice*. She proposes that an outcome should be designated in such a way that the agent's expectation of its value is the same under each event (Mahtani 2024, p.121).

The underlying rationale brings *Surgery* under the same diagnosis. 'M' designates the amount in the envelope you picked, and that amount is the smaller of the two at e_1 and the larger at e_2. 'Firsty' designates whoever arrives first, and that person is Ann at e_1 and Belle at e_2. In each case the table is filled in using a designator that does not hold its referent fixed across the events by which the table is partitioned. I therefore propose:

Heuristic 1 (Constancy across the events): a designator occurring in the outcomes of a decision table should pick out the same object at every event of that table.

Applied to *Two Envelope*, Heuristic 1 rejects Tables 8 and 9 alike and selects the table whose outcomes are designated by the smaller of the two amounts, which assigns

equal expected value to sticking and switching. Applied to Surgery, it rejects Table 2 and retains Table 1—which is the general form of the constraint of §6.1. Note its scope: it constrains which designators may fill the outcomes of a table whose events are *already* fixed. It says nothing about which partition supplies those events. That question is settled by fit, as §6.1 settled it.

There is reason to think the heuristic is not *ad hoc*. A prospect is a weighted sum of the utilities accruing to a single individual across the states, and an expected utility is a weighted sum of the values of a single outcome-variable across them. Summing requires that there be one individual, or one outcome-variable, whose fortunes at the several states are being aggregated. Heuristic 1 therefore follows from the nature of prospects. This is the same observation from which Gustafsson and Kowalczyk begin, when they liken a prospect to a probability-weighted portfolio of what can happen to a given person. They draw from it the conclusion that an identification method must be supplied from outside the case. I draw the conclusion that the designators used must hold their reference fixed across the events.

Three objections deserve replies.

The first is Mahtani's own. She considers focusing on designators that pick out the same person at every state, connects the proposal to rigidity, and objects that no principled distinction between rigid and non-rigid designators is available in the setting of epistemic possibility (Mahtani 2021, p.2087). Where the states are epistemic possibilities populated by distinct individuals, there are no trans-possibility individuals for a designator to track, and the distinction indeed has no application. Where the states are metaphysical possibilities, the distinction is the ordinary one and the heuristic is well defined. The availability of Heuristic 1 thus stands or falls with the reading of the states argued for in §6.2.

Mahtani also notes that a constraint of this shape can be evaded by redesignation, since a description may be rigidified—'the F' replaced by 'the actual F' (Mahtani 2021). Applied here, 'Firsty' would become 'the actual Firsty'. But rigidity and constancy across the events come apart. 'The actual Firsty' denotes whoever *in fact* arrives first, and either that denotation is held fixed to the world which is in fact actual, in which case the reasoner cannot fill in the table, since she does not know who arrives first; or it is evaluated at each event in turn, as filling in a column requires, in which case it denotes Ann at e_1 and Belle at e_2, and the heuristic applies.

The second objection concerns scope. In *Surgery* the two designator sets are conspicuously unalike, since 'Firsty' and 'Secondy' are stipulated descriptions while 'Ann' and 'Belle' are proper names. One might think the heuristic exploits an asymmetry that need not be present. Mahtani notes that gerrymandering is inessential to her cases (Mahtani 2021, p.2088). Gustafsson and Kowalczyk's version uses two perfectly ordinary pairs of names, 'Ann' and 'Beth' against 'Smith' and 'Jones'. The reply is that the asymmetry the heuristic exploits is not between kinds of designator but between

uses of them. If 'Smith' rigidly designates whichever patient in fact bears that surname, then it designates one woman at every state, the table using 'Smith' and 'Jones' is the table using 'Ann' and 'Beth' relabelled, and there is no second representation to conflict with the first. A second representation appears only if 'Smith' is used to pick out whoever bears the surname at the state in question; that is, descriptively—and then it is non-rigid across the events and the heuristic applies. The reasoner may of course continue to use the surnames; what she may *not* do is treat them as picking out different women at different events while summing utilities across those events.

The third objection is that the heuristic may forbid too much. Non-rigid designators are useful, and decision tables mentioning the winner of a race, or next year's exchange rate, are not thereby defective. The restriction is narrower than that. It applies to designators occurring in the outcomes of a table whose events their reference varies across, where the point of the outcomes is to record what accrues to a given individual or variable at each event. A table may perfectly well mention the winner of a race in the description of its events, or in an outcome whose value does not depend on who the winner is. Savage makes a related observation about the individuation of outcomes (Savage 1954, p.84).

Heuristic 1 does not extend to *Simpson*. No designator there varies across the events of Tables 6 and 7, since the variation lies in which partition supplies the events. Nor does it, on its own, single out a unique table even where it applies: it rules representations out, and several may survive it. The second heuristic addresses both limitations.

7.3 Heuristic 2: internalise the structural uncertainty

Heuristic 2: where a situation admits several candidate problem-specifications that differ over some structural fact—the identity-specifications in *Surgery*, the causal graph in *Simpson*—the reasoner should construct a single model whose state space *includes that fact as a variable*, distribute her credence over its values, and apply her decision rule once to the resulting model.

Why not advise instead that the reasoner compute a verdict under each candidate specification, and then combine the verdicts? Bayesian model averaging works because averaging predictive distributions is marginalisation over a model index (Hoeting et al. 1999), and marginalisation is defined for probabilities.

The reason is that verdicts do not admit the operation. The candidate specifications in *Surgery* deliver the claim that FCFS is *ex ante* Pareto superior and the claim that it is not; and there is no weighted average of a claim and its negation. Adjusting for C and adjusting for D in Simpson deliver 'treat' and 'do not treat'. Averaging is well defined at the level of the model and undefined at the level of its output, so the heuristic must operate on the former.

Internalising the uncertainty does not always yield a compromise. In *Surgery* it yields Table 1's verdict outright, for the reasons of §6.1. In *Simpson* it yields a genuine

mixture. This is because the candidate graphs disagree about the sign of the effect and the reasoner's credence is spread over them. The heuristic degrades gracefully, in that a structural hypothesis with very high probability dominates the resulting distribution; this desirable property is inherited from the underlying marginalisation.

Three objections again deserve replies.

The first is that Heuristic 2 has nothing to work on. The heuristic tells the reasoner to put the structural fact into her state space and spread credence over its values, which presupposes both that there *is* such a fact, and that she is ignorant of it. On the contextual view of §3.2 there is none. An identification method is selected by context rather than discovered, and one cannot straightforwardly have a credence as to which method is operative.

The objection restates the dilemma of §6.2, and it is met by taking the first horn. Where the states can be read as metaphysical possibilities there is a fact about who is who, the reasoner is ignorant of it, and Heuristic 2 applies as stated. Where they cannot, it does not apply, and the price of that is the disjunction of §6.3. Which reading to adopt is itself a modelling choice; and by Interpretivism, a model on which permissibility is a matter of fact is to be preferred, other things equal, to one on which it turns on salience.

The second objection is that indeterminacy would serve as well. Why should the reasoner internalise her uncertainty, rather than conclude that her situation determines no unique problem and that there is nothing she determinately ought to do? Indeterminacy is a coherent position.

Two things count against it here. It offers no guidance, and in both cases the reasoner must act: the dose must be given or divided, and Sophie must be treated or not. And it fits her worse. She does have credences over the structural hypotheses—equal credence in the two identity-specifications in *Surgery*, some distribution over causal graphs in *Simpson*—so a model representing her as facing no determinate problem discards information she has, which is the very failing §6.1 found in models built on a single designator set. Indeterminacy is the right response where she has no such credences, and that is a different case.

The third is that a regress threatens. Spreading credence over candidate causal graphs requires a prior over graphs, and choosing that prior is itself a specification choice. If every such choice must be made by internalising uncertainty about it, the reasoner never begins.

Three things limit the damage. First, this is the ordinary situation in hierarchical Bayesian modelling, where it is not usually thought vicious: posteriors are typically less sensitive to higher-level priors than to lower-level ones, so the practical consequences attenuate as one ascends. Second, at some level equivocation is defensible

on independent grounds, since the entropy-maximising choice minimises worst-case expected loss (Chapter 9, J. Williamson 2016). Third, and most importantly here, the regress is faced by every proposal in the vicinity. A reasoner who selects an identification method by consulting context must still determine which context she is in, and nothing in the contextual proposal makes that determination self-executing.

7.4 What the heuristics do not do

The heuristics are *constraints* on acceptable specifications rather than a procedure for generating a specification. Heuristic 1 rules representations out; Heuristic 2 tells the reasoner what to do with those that survive, given that she has credences over the relevant structural hypotheses. Neither tells her which variables belong in the model in the first place, which is the representation selection challenge in its general form, and I doubt that any short principle will.

They are also silent on granularity. It is intuitive that a problem-specification should flesh out the states of the world finely, lest the problem lose its bearing on reality; there is an analogue in the importance of specifying cases finely when reasoning from examples in philosophy (Williams 1973; Wilson 2016). It does not follow that specifications should be *maximally* fine. §4.1 has already given one reason why. A maximally fine state-partition uses the finest relevant partition variable to divide the states into homogeneous reference classes, and that variable may be a mediator or a collider in the true causal structure; specifying the problem with respect to it may then yield a badly misleading estimate of the target effect. A second reason is that a specification reflecting the target situation in very high detail may be solvable only by models that overfit to features unlikely to recur (§2.6, T. Williamson 2024). Heuristic 2 goes some way here, since it permits the reasoner to spread credence over specifications of differing granularity and propagate the result. It does not tell her which granularities are candidates.

8 Conclusion

Mahtani's puzzles appeared to devastate Bayesian epistemology and orthodox decision theory. They left us in a dilemma between the jeopardised Bayesian framework, and a tenet whose compatibility with that framework was in question. Gustafsson and Kowalczyk showed that, on certain assumptions, that dilemma can be avoided. They located the trouble in the identification of individuals across states.

I have argued instead that the trouble lies one level up. Puzzles of the same kind arise even where no individual's identity is in question at all. My *Simpson* case shows this. The containment of the opaque-identity puzzle leaves this problem untouched. Moreover, if the Partition Assignment Thesis is true, the two kinds of case have an identical root. Selecting a representation is a modelling choice, governed by which model best fits the reasoner's epistemic situation; leaving it to context forces a choice between a contested ethical commitment and salience-sensitive permissibility, and the approach I prefer, *EM*, avoids both. What remains is the far more general question of how

inchoate, real-world situations determine well-defined formal problems. The heuristics do not answer the specification challenge. They show what an answer can look like—a constraint on admissible specifications rather than a procedure, defended on grounds internal to modelling practice rather than imported from a theory of content.

References

Aloni, M. (2001). "Quantification under Conceptual Covers". PhD thesis. University of Amsterdam.

Bickel, P.J., E.A Hammel, and J.W O'Connell (1975). "Sex Bias in Graduate Admissions: Data from Berkeley". In: *Science* 187.4175, pp. 398–404. DOI: http://dx.doi.org/10.1126/science.187.4175.398.

Cartwright, N. (1979). "Causal Laws and Effective Strategies". In: *Noûs* 13.4, pp. 419–437.

Chalmers, D. (2011). "The Nature of Epistemic Space". In: *Epistemic Modality*. Ed. by A. Egan and B. Weatherson. Oxford University Press, pp. 60–107.

Evans, G. (1982). *The Varieties of Reference*. Oxford: Clarendon Press.

Gerbrandy, J. (2000). "Identity in Epistemic Semantics". In: *Logic, Language and Computation: Volume 3*. Ed. by L. Cavedon et al. Stanford, pp. 147–159.

Goodman, N. (1978). *Ways of Worldmaking*. Hassocks: Harvester Press.

Greco, D. (2023). *Idealization in Epistemology: A Modest Modeling Approach*. Oxford: Oxford University Press.

Gustafsson, J. E. and K. Kowalczyk (2026). "Ex-Ante Pareto and the Opaque-Identity Puzzle". In: *The Journal of Philosophy, Forthcoming* 123.1.

Hesslow, G. (1976). "Two Notes on the Probabilistic Approach to Causality". In: *Philosophy of Science* 43.2, pp. 290–292. DOI: http://dx.doi.org/10.1086/288684.

Hoeting, J. A. et al. (1999). "Bayesian model averaging: a tutorial". In: *Statistical Science* 14.4, pp. 382–401.

Jeffrey, R. (1965). *The Logic of Decision*. New York: McGraw-Hill.

Mahtani, A. (2017). "The Ex Ante Pareto Principle". In: *Journal of Philosophy* 114.6, pp. 303–323. DOI: http://dx.doi.org/10.5840/jphil2017114622.

— (2021). "Frege's puzzle and the ex ante Pareto principle". In: *Philosophical Studies* 178.6, pp. 2077–2100.

— (2024). *The Objects of Credence*. Oxford: Oxford University Press.

Ninan, D. (2018). "Quantification and Epistemic Modality". In: *The Philosophical Review* 127.4, pp. 433–485.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference: 2nd Edition*. Cambridge: Cambridge University Press.

Rothschild, D. (2022). "The Scope of Bayesianism". In: *Draft*.

Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.

Spirtes, P., C. Glymour, and R. Scheines (2000). *Causation, Prediction and Search, 2nd Edition*. Cambridge: MIT Press.

Thorn, P. (2017). "On the preference for more specific reference classes". In: *Synthese* 194.6, pp. 2025–2051.

Titelbaum, M. (2012). *Quitting Certainties: A Bayesian Framework Modeling Degrees of Belief*. Oxford University Press.

Wallmann, C. (2017). "A Bayesian Solution to the Conflict of Narrowness and Precision in Direct Inference". In: *Journal for general philosophy of science* 48.3, pp. 485–500.

Wallmann, C. and J. Williamson (2017). "Four approaches to the reference class problem". In: *Making it Formally Explicit: Probability, Causality and Indeterminism*. Ed. by G. Hofer-Szabó and L. Wroński. Springer, pp. 61–81.

Weisberg, J. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.

Williams, B. (1973). *Problems of the self : philosophical papers, 1956-1972*. Cambridge: Cambridge University Press.

Williamson, J. (2016). *Lectures on Inductive Logic*. Oxford: Oxford University Press.

Williamson, T. (2021). "Epistemological Consequences of Frege Puzzles". In: *Philosophical topics* 49.2, pp. 287–319.

— (2024). *Overfitting and Heuristics in Philosophy*. New York: Oxford University Press.

Wilson, J. (2016). "Internal and External Validity in Thought Experiments". In: *Proceedings of the Aristotelian Society* 116.2, pp. 127–152.