Model behaviour: Mahtani's puzzles as problems of model-selection

Harry Long · University College London

1. SURGERY, AND THE PUZZLE

Surgery. Ann and Belle, same condition, arriving within the hour; you have no idea who arrives first and think it equally likely that each does. One dose. Either give a full dose to whoever arrives first (FCFS) or split it. A full dose is much more effective: better a one-in-two chance of a full dose than a half dose for sure. (Mahtani 2017)

Table 1. Using {'Ann', 'Belle'}; cells are Ann / Belle

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

FCFS *is ex ante* Pareto superior to Split.

Table 2. Using {'Firsty', 'Secondy'}; cells are F / S

	e_1: Ann first	e_2: Belle first	Prospects
FCFS	10 / 0	10 / 0	10 / 0
Split	4 / 4	4 / 4	4 / 4

FCFS is *not ex ante* Pareto superior to Split.

Prospect. 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.

Ex ante Pareto superiority. An action is *ex ante* Pareto superior to another iff 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. If an action is *ex ante* Pareto superior to another, then you should not do the latter.

Admissible designator set. A set of designators is admissible iff there is a bijective denotation function from that set to the set of individuals affected by one's actions.

Identity-specifications. {(Ann is Firsty and Belle is Secondy), (Ann is Secondy and Belle is Firsty)}. The reasoner has both admissible sets and cannot match their referents up.

Mahtani puzzle. 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.

2. TWO EXISTING DIAGNOSES

The tenet (Mahtani 2024). 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.

– So the reasoner cannot assign a credence to *that individual arrives first* independently of how the individual is designated. Prospects, and so facts about superiority, inherit the relativity.
– *Cost*: no account of the states in Ω compatible with the tenet is forthcoming. Possible worlds ⇒ objects of credence unclear; sentence-worlds ⇒ ambiguity, artificial language.

Trans-world identification (G&K 2026). A prospect specifies what happens to *that person* across the states, portfolio-style, so computing one presupposes a method of identifying individuals across states. Two are available (first name + condition; surname + arrival time) and the case singles neither out. Prospects are thus already relative to identification methods, and context supplies the method via a counterpart relation.

The transitivity argument. Suppose you know Smith arrives first. Trans-world identities cannot be read off the names.

(i) The case requires an epistemic possibility w_1 at which Ann is Smith, and one w_2 at which Ann and Smith are distinct.
(ii) Suppose we identify across the two by name. The individual at w_1 is Ann, hence identical with the referent of 'Ann' at w_2; she is also Smith, hence identical with the referent of 'Smith' at w_2.
(iii) But those two individuals at w_2 are distinct, and identity is transitive; and by symmetry there is no ground for identifying the w_1 individual with one rather than the other.
∴ Objects are not in general re-identifiable across epistemic possibilities; properties are (cf. Chalmers 2011).

What they share. Both locate the relativity in the semantics or metaphysics of prospect-ascription, and both then face a *selection question* (which designator set? which identification?) that neither treats as belonging to the theory of prospects. **My proposal**: the selection question is the whole of the problem, and it is a question about models.

3. SIMPSON: A MAHTANI PUZZLE WITH NO IDENTITY UNCERTAINTY

Simpson. Frequencies, your credences and the objective conditional probabilities coincide. Your evidence is the data below. You know which reference class each affected individual falls in for every relevant partition variable. You do *not* know the causal structure. Sophie has small stones (c) and is under fifty (d). Treat (b) or not (b')? $U(a) = 1$, $U(a') = 0$.

Tables 3–5. Study data (N = 105,500)

	by B		by B and C				by B and D			
	b	b'	bc	$b'c$	bc'	$b'c'$	bd	$b'd$	bd'	$b'd'$
Success (a)	5,475	25,250	5,000	250	475	25,000	25	500	5,450	24,750
Failure (a')	45,025	29,750	45,000	4,750	25	25,000	475	4,500	44,550	25,250
Total	50,500	55,000	50,000	5,000	500	50,000	500	5,000	50,000	50,000

Derived conditionals, and the reversal

class		$P(a \mid b\cdot)$	$P(a \mid b'\cdot)$	treatment
overall	(1)	.108	.459	worse
small stones (c)	(2)	.100	.050	**better**
large stones (c')	(3)	.950	.500	**better**
under 50 (d)	(6)	.050	.100	worse
50 or over (d')	(7)	.109	.495	worse

Tables 6–7. The decision, $\{a, a'\}$ as events

Sophie placed in	EU(treat)	EU(don't)	verdict
the c-class (T6)	0.10	0.05	treat
the d-class (T7)	0.05	0.10	don't

Why not the finest class? (i) Margins do not fix the joint: $P(a|bcd) \in [0, .053]$ and $P(a|b'cd) \in [0, 1]$, so the sign of the difference is open. (ii) The finest partition variable may be a mediator or collider, and so not one to condition on — and the causal structure is unknown.

Engaging the principle under containment. Add **Tom**: large stones (c'), under fifty (d). Policies: treat both, or treat neither.

– By stone size: treatment better for Sophie (.100 vs .050) *and* for Tom (.950 vs .500) ⇒ **treat both** is *ex ante* Pareto superior.
– By age: both are in d, and no treatment is better for each (.050 vs .100) ⇒ **treat neither** is *ex ante* Pareto superior.

A claim and its negation, with no identity uncertainty anywhere. G&K's containment argument works by locating the source in the epistemic/metaphysical distinction, and so has no purchase here.

4. A COMMON ROOT

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.

Let ρ = the probability that the individual to whom a prospect is assigned arrives first. Relative to {'Ann', 'Belle'}, $\rho = .5$ for each (Table 1); relative to {'Firsty', 'Secondy'}, ρ is 1 for one and 0 for the other (Table 2). Sophie: P(survives | treated) is .10 relative to the assignment by stone size, .05 relative to that by age. In each case a total probability is well defined only once the individual is assigned to a cell, and the evidence leaves the assignment open. *Support*: descriptivism about designators; Aloni's conceptual covers; G&K's own identification methods, which are given by properties.

5. EPISTEMOLOGY AS MODELLING

- ***Unavoidability.*** To talk of beliefs, knowledge or credences is already to model within a framework; there is no doing the former without the latter.
- ***Interpretivism.*** For S to have credence n that p in C is for the best Bayesian model of S in C to represent her so; where the best model is that with maximal simplicity, strength, and fit.
- ***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.*** Absent a perfect model of the domain, we may use different models in different contexts.

Surgery on EM. A model built on either designator set alone represents her as knowing which individuals it picks out, and fails on *fit*. Which set is held fixed is settled by fit too: she knows which patient is Ann; she does not know whether Ann arrives first.

G&K's dilemma. Both the transitivity argument and Mahtani's general argument need some designator to shift its reference across the states. Neither is forced here, since the uncertainty is about *contingent* matters (who arrives first; who bears which surname). So: **either** the states read as metaphysical possibilities, in which case the identification is supplied by the case and the counterpart apparatus is idle; **or** they do not, in which case the apparatus is needed and brings the cost below. I take the first horn.

Cost of the second horn for G&K. Identifications redistribute the same total expected welfare among persons: FCFS gives $5+5$ on one, $10+0$ on the other — ten either way; Split gives eight either way. So a criterion ranking options by the *total* is invariant across identifications, and any criterion sensitive to how the total is *distributed* must vary. Hence: either accept expectational utilitarianism, or accept that permissibility can turn on what is merely salient.

6. THE SPECIFICATION CHALLENGE, AND TWO HEURISTICS

The specification challenge. To describe 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. Two aspects: *representation selection* and *granularity*.

Tables 8–9. *Two Envelope*

	e_1: $N = 2M$	e_2: $N = 0.5M$	EU
Stick	M	M	M
Switch	$2M$	$0.5M$	$1.25M$
Stick	$0.5N$	$2N$	$1.25N$
Switch	N	N	N

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.

Scope: constrains which designators may fill the outcomes of a table whose events are *already* fixed; silent on which partition supplies the events. That is settled by fit (§5).

Heuristic 2 (Internalise the structural uncertainty). 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.

What the heuristics do not do. They are constraints on admissible specifications, not a procedure for generating one.

7. OBJECTIONS AND REPLIES

- **Objection.** There is no principled distinction between rigid and non-rigid designators in the setting of epistemic possibility, so Heuristic 1 has no content (Mahtani 2021, p.2087). **Reply.** Granted, *where* the states are epistemic possibilities populated by distinct individuals: there is then no trans-possibility individual for a designator to track. Where they are metaphysical possibilities the distinction is the ordinary one. So Heuristic 1 stands or falls with the reading of the states defended in §5—the first horn of the dilemma, and the horn I take.
- **Objection.** The constraint can be evaded by redesignation: replace 'the F' with 'the actual F', which *is* rigid. A constraint stated in terms of rigidity therefore excludes nothing (Mahtani 2021, p.2083). **Reply.** Rigidity and constancy across the events are different properties, and Heuristic 1 requires the second. 'The actual Firsty' denotes whoever *in fact* arrives first. Either that denotation is held fixed to the world which is in fact actual—and then 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, and then it denotes Ann at e_1 and Belle at e_2. Rigidifying fixes the world of evaluation; the events of a table are candidates for actuality.
- **Objection.** Heuristic 1 exploits an asymmetry that need not be present. In *Surgery* one set is stipulated descriptions and the other proper names, but G&K's version uses two ordinary pairs of names. **Reply.** The asymmetry is between *uses* of designators, not kinds of them. If 'Smith' rigidly designates whichever patient in fact bears that surname, it designates one woman at every event, and the surname table is the first-name table relabelled: no second representation arises. A second representation arises only on the *descriptive* use—whoever bears the surname at the event in question—and then the reference varies across the events and the heuristic applies.
- **Objection.** Heuristic 1 forbids too much: tables mentioning the winner of a race, or next year's exchange rate, are not thereby defective. **Reply.** It applies only to designators in the *outcomes* of a table across whose events their reference varies, where the outcomes record what accrues to one individual at each event. A table may mention the winner of a race in its *events* (cf. Savage 1954, p.84).
- **Objection.** *EM* is itself context-relative. Specification Interpretivism makes the problem relative to a context C, so a shift in salience is a shift in C, and permissibility turns on salience here too. **Reply.** It depends what fixes C: here, the reasoner's evidence and the decision before her, not what is conversationally salient. Two of G&K's own variants separate the views. A patient telephones and rings off before giving her first name: nothing is learnt about who is who, and *EM*'s model does not shift. A dossier on Smith and Jones arrives, full of detail but no first names: something *is* learnt, and *EM*'s model may shift. A counterpart relation fixed by salient resemblance shifts in both alike, and so cannot mark the difference.
- **Objection.** Why adopt a metaepistemology at all? G&K's solution needs none. **Reply.** Against Mahtani, *EM* is the cheaper option: the tenet requires an account of the objects of credence that she herself shows to be unavailable. Against G&K what the commitment to *EM* buys is generality. Their answer is an answer about identifying individuals across states, and is silent where no individual's identity is in question—as in *Simpson*.

Key references. Aloni (2001). Chalmers, 'The Nature of Epistemic Space' (2011). Greco, *Idealization in Epistemology* (2023). Gustafsson & Kowalczyk, 'Ex-Ante Pareto and the Opaque-Identity Puzzle', *JPhil* 123.1 (2026). Mahtani, *JPhil* (2017); *Phil. Studies* (2021); *The Objects of Credence* (2024). Pearl (2009). Savage (1954). Titelbaum (2012). T. Williamson, *Phil. Topics* (2021); *Overfitting and Heuristics* (2024).