Diagonal Decision Theory

Melissa Fusco

A natural view from probability theory is that the objects of credence functions, like the objects of any probability function, are propositions. In her book *The Objects of Credence*, however, Anna Mahtani reminds us that this view gives rise to a number of familiar philosophical puzzles. Suppose you point at Carnap and tell me: "he is German". I don't know the man you're pointing at *is* Carnap, however. It seems that I can rationally assign different credences to

(1) He is German.

(2) Carnap is German.

so that (1) and (2) provide me with *different* objects of credence, despite expressing the *same* proposition.

Such examples also raise questions about the objects of *learning*. Since it seems that what I *learned* from you is (1) and not (2), the objects of learning appear to be more sentence-like than proposition-like.

Stalnaker's "Assertion" (1978) offers a classic analysis of the second puzzle: that is, to the question of how rational agents should update in cases where the referents of context-sensitive terms like *here*, *now*, *he*, and *I* differ in different epistemically possible worlds. Stalnaker's answer involves a move known as *diagonalization*. Here, I explore the relationship between diagonalization and some puzzles in Mahtani's book. We'll look at its influence on two of Mahtani's case-studies, the Principal Principle (Ch. 5) and the finite version of the Two Envelope Paradox (Ch. 6).

1 Introducing the Diagonal

I am trying to keep the peace at a family reunion where everyone has different political views. The family goes to a flea market, where we come across a framed picture for sale. I inspect it and realize it is a rendering of actor Alec Baldwin lampooning Donald Trump on Saturday Night Live. I point to the picture and say:

(3) He makes me smile.

Three people overhear me: my conservative uncle, my cousin, and my grandfather.

My conservative uncle looks at the picture and forms the belief that it is of Trump himself. My uncle also believes that I and everyone else in the family are big Trump fans. He concludes on this basis that what I said is true. Let w_1 be the world my uncle thinks we are in.

My cousin takes a look at the picture and forms the (true) belief that it is Baldwin in the picture. He believes that I enjoy watching Trump lampooned. So he agrees with my uncle that what I said was true—though he disagrees with my uncle about the *content* of what I said. (As a heuristic, we can think of the possible contents of sentences like (3) in terms of their disquoted analogues with the context-sensitive terms removed. So to my uncle, the content of what I said with (3) is *Donald Trump makes MF smile*; to my cousin, it is *Alec Baldwin makes MF smile*.) Let w_2 be the world my cousin thinks we are in.

Finally, my grandfather takes a look at the picture and forms the belief that it depicts Juan Trippe, the 1930s aviation pioneer.[1] Knowing I have no particular knowledge or affection for *that* man, my grandfather concludes that what I said is false. His is yet a third nonequivalent hypothesis about the content of my utterance, viz., *Juan Trippe makes MF smile*. Let w_3 be the world my grandfather thinks we're in.

As Stalnaker puts it, such examples are of interest because they highlight two ways the truth-value of my utterance depends on the way the world is (pg. 78). The first way is a matter of reference: what (a particular token of) "he" refers to. The second way is a matter of worldly facts. Here, those worldly facts concern: who in the world, independent of any facts about the *in situ* referents of actual and counterfactual pronouns, makes me smile.

Figure 1 depicts a two-dimensional semantic matrix for sentence (3), showing the different truth-values the sentence takes with respect to every pair $\langle w_i, w_j \rangle$ of worlds drawn from $\{w_1, w_2, w_3\}$.

	w_1	w_2	w_3
w_1	T	F	F
w_2	F	T	T
w_3	F	F	F

Figure 1: Matrix for (3).

In the matrix, worlds appear on the vertical axis in their role as *world-as-actual*, determining the content expressed by the sentence. At the level of the whole sentence, we can think of this as determining the proposition expressed by (3): as we saw, the candidates are, in order: {*Donald Trump makes MF smile, Alec Baldwin makes MF smile, Juan Trippe makes MF smile*}. Worlds in the matrix appear in the horizontal axis in their role as *world of evaluation*, where the truth of that proposition is interrogated. Whether a T ("true") or an F ("false") appears here often depends on counterfactual information of great conjectural interest to philosophers.[2] For example, on the bottom left, at $\langle w_3, w_1 \rangle$, we see an F: the claim here is that proposition *Juan Trippe makes MF smile* is false in the world my uncle thinks we're in—which is to say: in a world where I and my family are all big Trump fans and Trump's picture crops up at the flea market, I (still) don't have a special place in my heart for figures in the history of commercial aviation. That I lack such affection in w_1 is not strictly entailed by the story I told at the beginning. It is a (reasonable) counterfactual stipulation.

Suppose we want to put all such counterfactual information aside, and summarize only the facts about whether my utterance is consistent with my family members' views on what the *actual* world is like. The diagonal of the matrix (highlighted in Figure 2) brings out just this: my uncle thinks what I said was true (T), as does my cousin, but my grandfather thinks what I said was false (F). A third party genuinely uncertain about which of the worlds in $\{w_1, w_2, w_3\}$ is actual is still in a position to see that my utterance is consistent with exactly the first and second of these worlds.

We *diagonalize* the matrix when we apply the dagger (†) operator of 2D semantics to (3) (Figure 2 right), "projecting" just this information across every row.

[1] Trippe was, in fact, portrayed by Baldwin in the 2004 film *The Aviator*.

[2] For example, whether the proposition expressed by a sentence is metaphysically necessary will depend on this counterfactual profile (Kripke, 1980).

	w_1	w_2	w_3
w_1	T	F	F
w_2	F	T	T
w_3	F	F	F

	w_1	w_2	w_3
w_1	T	T	F
w_2	T	T	F
w_3	T	T	F

Figure 2: The matrix for (3) (left), and for †(3) (right)

This operation captures Stalnaker's answer to the question we began with: how a rational agent should update in cases where the referents of context-sensitive terms (here, *he*) differ across epistemically possible worlds.[3] His conjecture is that no matter which world is actual, the rational thing to do is keep in play exactly the worlds consistent with the diagonalized matrix. If a bystander who starts out having no idea whether we're in world w_1, w_2, or w_3 hears me utter sentence (3), and believes I am honest, he should "throw out" w_3, but keep w_1 and w_2.

2 2D Thirding in Sleeping Beauty

Let's apply diagonalization to a puzzle of credal self-location: the Sleeping Beauty problem discussed by Elga (2000). On Sunday, Beauty signs up for a laboratory experiment. First, she'll be put to sleep, and a fair coin will be tossed. Whether the coin comes up heads or tails, Beauty will be awakened briefly on Monday. Moreover, if the coin comes up heads, she'll sleep through Tuesday; otherwise, she'll be woken again briefly on Tuesday (Figure 4). Each time she awakens, a drug will erase her memory of any previous wakings.

	Monday	Tuesday
heads	awake w_1	asleep w_4
tails	awake w_2	awake w_3

Figure 3: Sleeping Beauty Experiment.

The next thing Beauty knows, she wakes up in the lab. The experimenter asks her a question:

(SB) What's your credence the coin landed heads?

In Sleeping Beauty, there are two intuitively rational answers to (SB). The first is 1/2: the coin is fair. The second answer is 1/3. This answer is suggested by Figure 3: of the three cells of the table consistent with Beauty's being awake, only one of them coincides with the coin's landing heads. The challenge for this answer is to articulate a reply to a subsequent question: given that Beauty *started out* with credence 1/2 in the coin's landing heads, what proposition p did she learn upon waking up, such that conditioning her prior on p results in a posterior of 1/3?

Diagonalization can help, by increasing the expressive capacity of the language relative to Beauty's

[3] The meanings of these terms will be familiar readers from the semantics literature (Kaplan, 1989):

term	referent, relative to context c
"here"	the place of c
"you"	the addressee of c
"I"/"me"	the speaker of c
"now"	the moment of c
"the actual F"	the unique F at w_c, the world of the context
"she"	the salient female in c

epistemically possible worlds. Begin with a sentence Beauty can truly utter when she wakes up:

(4) I am awake today.

As we saw in the Trump-Baldwin example, sentence (4), which contains the indexical expression "today", expresses different propositions at different worlds in Beauty's information state (Figure 4).[4]

at	the day is ...	(4) expresses ...
w_1	Monday	$\{w :$ Beauty awakens on Monday in $w\}$
w_2	Monday	$\{w :$ Beauty awakens on Monday in $w\}$
w_3	Tuesday	$\{w :$ Beauty awakens on Tuesday in $w\}$
w_4	Tuesday	$\{w :$ Beauty awakens on Tuesday in $w\}$

Figure 4: Candidate propositions for (4).

In some worlds (w_1 and w_2), (4) expresses the *a priori* proposition *Beauty awakens on Monday*—*a priori* in the sense that Beauty knew all along that she would wake up on Monday. But in worlds w_3 and w_4, (4) expresses the previously-uncertain, chance-.5 proposition *Beauty awakens on Tuesday*—chance .5 because this was slated to happen iff the coin came up tails (viz., as per Figure 4, only in worlds w_2 and w_3).

It is just this predicament, for Stalnaker, that rationally calls for diagonalization. Figure 5 shows the semantic matrix for (4), with the four possible worlds in Figure 4 arrayed along each axis. The diagonal is highlighted.

	w_1	w_2	w_3	w_4
w_1	T	T	T	T
w_2	T	T	T	T
w_3	F	T	T	F
w_4	F	T	T	F

Figure 5: *Heads* is true at 1/3 of †(4)-worlds along the diagonal.

Leveraging the assumption that, upon waking, Beauty's credence function is conditioned on the proposition expressed by the diagonal of the matrix—viz., on †(4)—we can model the thirder answer: $Cr(\text{heads} \mid \dagger(4)) = 1/3$. The total package involves an appeal to Conditionalization enriched with a two-dimensional bridge principle; we can call the combined norm "Sentential Conditionalization":

> **Conditionalization:** If a rational agent with prior $Cr(\cdot)$ learns proposition E, her posterior is $Cr(\cdot \mid E)$.

> **Sentential Conditionalization:** If a rational agent with prior $Cr(\cdot)$ learns an evidence-sentence A, her posterior is $Cr(\cdot \mid \dagger A)$.

The Equivalence Lemma of two-dimensional logic—which roughly states that diagonalization is inert for sentences which do not contain "disputed" context-sensitive terms[5]—allows us to write the

[4] I use the "world(s)" terminology here because of its match with the literature, though readers will note that $w_1 - w_4$ are really world-time pairs (w_1: $\langle$ the heads world, Monday $\rangle$; w_2: $\langle$ the tails-world, Monday $\rangle$ and so forth.)

[5] Stalnaker *op. cit.*, pg. 82. The underlying fact about two-dimensional logic is that the dagger operator † is idempotent: $\dagger S = \dagger\dagger S$.

new Conditionalization norm with daggers scoped over *both* the evidence-sentences and hypothesis sentences:

> **Conditionalization, II:** If a rational agent with prior $Cr(\cdot)$ learns proposition E, her posterior in H is $Cr(H \mid E)$.

> **Sentential Conditionalization, II:** If a rational agent with prior $Cr(\cdot)$ learns evidence-sentence A, her posterior in sentence B is $Cr(\dagger B \mid \dagger A)$.

So in Beauty's case: $Cr(\text{heads} \mid \dagger(4)) = Cr(\dagger\text{heads} \mid \dagger(4)) = 1/3$.

To model the appeal of the Stalnakerian analysis, it is helpful to reflect on what happens if we *don't* diagonalize. Suppose one maintains that, when one receives testimony S at world w_i, one is rationally required to update on *the proposition S expresses at w_i*, rather than on $\dagger S$ (which, as we noted, expresses the same proposition relative to any world in w_j in the support of Cr). Such insistence is vulnerable to **an objection from stringency:** un-diagonalized norms generate unreasonable demands on rational agents. Suppose, unbeknownst to her, that Beauty has in fact woken up on Tuesday. We can see from Figure 4 that she is therefore in w_3. Should she be rationally required to adopt $Cr(\text{heads}) = 0$? Intuitively, no, although this is $Cr(\text{heads} \mid (4))$ at the context she occupies.[6] The diagonalizer's answer $Cr(\text{heads} \mid \dagger(4))$ is much more natural.

Stepping back, it's now possible to be a bit more precise about how natural language is hypothesized to relate to the guiding question of Mahtani's book. Diagonalization lets us reply to the question:

(OC) What are the objects of credence?

with aspects of both the traditional answers ("sentences!", "propositions!"). An agent's credences can be modeled as distributed over a set of worlds she considers epistemically possible. Linguistic competence converts *sentences* to matrix semantic values, where these worlds to populate the vertical and horizontal axes. The diagonals of these matrices are *propositions*. It is *those propositions* which, at a first pass, are the objects of credence and update.

3 Old Norms in New Bottles

Now that diagonalization is up and running, we can consider diagonalized versions of other rational norms. A recent one, due to Dmitri Gallow, two-dimensionalizes Lewis's famous "Principal Principle". According to the Principal Principle, one's credence in a proposition H should defer to the objective chance (Ch) of H (Lewis, 1971; Mahtani Ch. 5.3). Gallow's sentential version, adapted to the present notation, is:

- **Sentential PP.** A rational agent's credence in ϕ should defer to her expectation of the chance of $\dagger\phi$ (Gallow, 2023, §3).

  $$Cr(\dagger\phi \mid Ch(\dagger\phi) = n) = n$$

To Sentential PP and Sentential Conditionalization, I'd propose to add a third. The rough idea is:

- **Sentential Utility.** The value of ϕ, for a rational agent, is the value of $\dagger\phi$.[7]

[6] As per Figure 6: at w_3, "I am awake now" expresses the proposition *tails*—viz., $\{w_2, w_3\}$—at w_3. *heads* and *tails* are mutually exclusive, so if Beauty learns $\{w_2, w_3\}$, she is rationally required to have credence zero in *heads* ($\{w_1, w_4\}$).

[7] I assume utilities are values in $\mathbb{R}$, ignoring the equivalences brought about by finite affine transformations.

In the next section, I'll discuss **Sentential Utility** in the context of Mahtani's finite version of the Two-Envelope Paradox (Ch. 6). First, a warm-up example should help us move from matrix semantic values (with range {T, F}—below left) to matrix values with range ℕ—below right.

	w_1	w_2	w_3	w_4
w_1	T	T	F	F
w_2	T	T	F	F
w_3	T	F	T	F
w_4	F	T	T	T

	w_1	w_2	w_3	w_4
w_1	10	20	30	40
w_2	10	20	30	40
w_3	20	40	60	80
w_4	20	40	60	80

Figure 6: Value matrices for a sentence ϕ

At first pass, it may seem silly to worry about functions from sentences to numbers of this kind. But—as I think Mahtani's Two Envelopes show—it isn't. Agents who do not use language cannot, for better or worse, enter into a wide range of betting arrangements, for the simple reason that the payoff conditions of such arrangements cannot plausibly be described to them. The question of *how* linguistic competence enters into the calculation of expected utility, when the outcomes are described with all the richness of natural language, is important and far from trivial.

Our warm-up will consider this sentence:

(5) I have as many Tamagotchi as I actually have.

What is—not the probability, but—the *value* of (5)?[8]

We consider the question relative to an information state with four possible worlds, w_1 through w_4. As in the Trump-Baldwin story, we need to fill in the nature of w_1 through w_4 in a away that makes vivid how the value of the outcome described by (5) depends both on the nonlinguistic facts and on the semantic facts about indexicals (here, "actually").

Here's a story that fits the bill. Say I have inherited a large treasure-chest from my grandmother, the Dread Pirate Roberta. I am digging through the coins and goblets inside when I find a Tamagotchi. How much is it worth? I do not know, and different worlds in my information state will represent different resolutions of my ignorance. Let's say there are four price possibilities: in w_1, a Tamagotchi is worth \$10; in w_2, a Tamagotchi is worth \$20; in w_3, \$30; and in w_4, \$40. All these worlds (and these alone) are epistemic possibilities for me.

In addition to my ignorance regarding the resale price of Tamagotchi, I am ignorant of how many Tamagotchi I have inherited. Perhaps there is an additional Tamagotchi in the box, and perhaps not; thus I do not know what $n \in \{1, 2\}$ the noun phrase "the actual number of Tamagotchi I have" denotes.

Finally, let us stipulate that, for some reason, these aspects of the world—the price of Tamagotchi, and the number of them I inherited—are not independent. Perhaps I know, given her hoarding habits, that the Dread Pirate would only have stashed a second Tamagotchi in the chest if each was individually quite valuable.

This example would then give us the right matrix in Figure 6 as the *utility value* (rather than the *semantic value*) of sentence (5). Once again, we have a case of **horizontal mismatch**; $Val(\dagger(5))$ is not equivalent to any un-diagonalized candidate for $Val(5)$. Rather, the diagonal represents a kind of value polarization: in worlds where I have fewer Tamagotchi, they are each individually worth less; in worlds where I have more (viz., two) Tamagotchi, each is individually worth more. **Sentential Utility** says: the value of the state of affairs expressed by (5) is the function from worlds to utilities expressed by the diagonal of the matrix.[9] Action-guiding norms that employ Val-expressions, such as expected

[8] The *probability* of (5) is surely 1!

[9] An off-diagonal value, like the '4' in the top-right of the matrix, again represents a *counterfactual* value: it is how

utility:

$$EU(A) = \sum_{E} Cr(E)Val(A \wedge E) \tag{1}$$

are targets for diagonalization because Val appears in the weighted sum. **Sentential Utility** encourages us to replace Equation (1) with Equation (2):

$$EU(A) = \sum_{E} Cr(\dagger E)Val(\dagger(A \wedge E)) \tag{2}$$

4 Making Value 2D in the Finite Two-Envelope Paradox

So much for the warm-up exercise. We turn to Mahtani's example, the Finite Two-Envelope Paradox.[10]

> Suppose that you have before you two envelopes, each of which contains a cheque...You have no idea how much money these two envelopes contain, but you do know that one contains twice as much money as the other. You select an envelope at random. But then you consider: should you stick with the envelope you've selected, or should you switch[?] There seems to be a good reason to switch, as you can see from the decision table below, with 'M' standing for the amount of money in the envelope in your hand, and 'N' standing for the amount of money in the other envelope:

	Event e_1: N = 2M Cr(e_1) = 0.5	Event e_2: N=0.5M Cr(e_2) = 0. 5	Expected Utility
Stick	M	M	M
Switch	2M	0.5M	1.25M

Table 6

Figure 7: Mahtani, pg. 117

> Obviously the expected utility of sticking with your current envelope is M (you are guaranteed to get the amount that is in that envelope); whereas if you switch there is a 0.5 chance that you will get double M, and a 0.5 chance that you will get half of M, and so the expected utility of switching is $(0.5)(2M) + (0.5)(0.5M) = 1.25M$. Switching then seems to have higher expected utility than sticking, and so [expected utility maximization] requires you to switch. But this is a very strange result!

She continues:

> And it is not just intuition that speaks against this result, for we can construct a parallel argument in favour of sticking, with the outcomes stated in terms of N (the amount of money in the other envelope)... What has gone wrong here?

much the number of Tamagotchi I have in w_1 (viz., one) would be worth in a world where each individual Tamagotchi is worth $40—even though I think that combination is ruled out by my information.

[10] This is a *finite* case. The resolution in the infinite case (seen on YouTube here) is taken to turn on the nature of sums which are merely conditionally convergent, which they can only be in a (countably) infinite case. So it's important that in Mahtani's version there are only *two* events, E_1 and E_2; explanations that exploit conditional convergence do not apply. I suspect—though I won't make the case here—that part of the reason for the large literature on the paradox is that *both* phenomena are involved.

As in (5), a two-dimensionalist will insert a neon "actually" in the descriptions M and N in Mahtani's vignette. We should start by glossing M and N as full sentences.

(6) *(Stick):* I get the envelope I actually have.
M

(7) *(Switch):* I get the envelope I don't actually have.
N

There are also two states:

(8) I have the lesser envelope.
E_1

(9) I have the greater envelope.
E_2

If we let x be the amount in the lesser envelope, we can rewrite Mahtani's Table 6 (with some convenient world-labels) as follows:

	I have lesser (E_1)	I have greater (E_2)
stick (M)	x w_1	$2x$ w_2
switch (N)	$2x$ w_3	x w_4

When we generate the matrices for M and N with the four worlds generated by $\{M, N\} \times \{E_1, E_2\}$, we see that the two acts denote different constant functions $W \to \mathbb{R}$ in different worlds (w_1 and w_2 are relevant for M; w_3 and w_4 for N).

	w_1	w_2	w_3	w_4
w_1	x	x	x	x
w_2	$2x$	$2x$	$2x$	$2x$
w_3	$2x$	$2x$	$2x$	$2x$
w_4	x	x	x	x

	w_1	w_2	w_3	w_4
w_1	$2x$	$2x$	$2x$	$2x$
w_2	x	x	x	x
w_3	x	x	x	x
w_4	$2x$	$2x$	$2x$	$2x$

Figure 8: Value of M (left) and N (right)

We can, therefore, tap the diagonalization strategy. It is easy to show, using Equation (2) rather than Equation (1), that the expected values of the two acts are in fact the same:

$$\begin{aligned} EU(M) &= \sum_E Cr(E)Val(\dagger(M \wedge E)) \\ &= Cr(E_1)Val(\dagger(M \wedge E_1)) + Cr(E_2)Val(\dagger(M \wedge E_2)) \\ &= .5 \cdot Val(\{w_1\}) + .5 \cdot Val(\{w_2\}) \\ &= .5 \cdot x + .5 \cdot 2x \\ &= 1.5 \cdot x \end{aligned}$$

$$
\begin{aligned}
EU(N) &= \sum_{E} Cr(E \mid N) Val(\dagger(N \wedge E)) \\
&= Cr(E_1) Val(\dagger(N \wedge E_1)) + Cr(E_2) Val(\dagger(N \wedge E_2)) \\
&= .5 \cdot Val(\{w_3\}) + .5 \cdot Val(\{w_4\}) \\
&= .5 \cdot x + .5 \cdot 2x \\
&= 1.5 \cdot x \quad \checkmark
\end{aligned}
$$

5 The Big Picture

I've argued that the two-dimensionalist's diagonalizing strategy has something to contribute to (at least) two of the puzzles studied by Mahtani. Here I briefly compare her own analysis. Two discussions in the book are particularly relevant. First, in the Two-Envelope case, Mahtani endorses a refinement of a strategy suggested by Horgan (2002):

> **Sameness of expectation of value:** An outcome in a decision table should be designated in such a way that under each event, the agent's expectation of the value of the outcome (so designated) is the same. (Mahtani, pg. 127)

Because Mahtani's decision table 6 (in Figure 8) violates the principle of **Sameness of expectation in value**, arguments based on it (like the argument that one should switch) are "defective" (127), and should be rejected.

Second and more generally, in Chapter 7 Mahtani critically discusses David Chalmers's two-dimensional approach to understanding credal reports (Chalmers, 2011a,b). In Chalmers's terminology—which is importantly related to Stalnaker's, though it connotes a different array of philosophical commitments—the objects of credence should be interpreted as *primary*, rather than *secondary*, intensions. That distinction maps roughly onto our claim that the objects of credence (and learning! and value!) are *diagonal*, rather than *horizontal*, propositions.

I begin with Horgan, whose recommendation is that outcomes must be stated in terms of "epistemically rigid designators" (qtd. Mahtani, pg. 127). What this comes to, I think, is that terms like M shouldn't be used in the formulation of decision problems. My *worry* about the strategy is how restrictive it is. For example, in Sleeping Beauty, "today" is not an epistemically rigid designator, because when Beauty wakes up, she doesn't know what day it is (*ergo*: does not know which day "today" denotes). Should we therefore prohibit Beauty from thinking about her situation using the concept (that she would articulate with) "today"? This move seems to deprive Beauty of *any* way of thinking about where she is located in time—to say nothing of precluding the attractive thirder strategy I sketched above.

When it comes to Chalmers's two-dimensionalist view, Mahtani worries about the destructive effects of a wholesale move from secondary to primary intensions. In §7.6, she lists three important ones: (i) thinking-alike, (ii) deference, and (iii) conditionalization:

> [(i)] The source of the problems is that our new convention disrupts a natural assumption made by users of the credence framework: that the objects of credence are common property. We are used to assuming that two different people can have credences in the very same object. . . on [Chalmers's] proposed convention it is hard to tell whether two credence attribution statements relate to the same object of credence. . . [(ii)] [A]nother. . . problem concerns deference principles. To illustrate, suppose that Tom regards Tess (so designated) as an expert. Then—given the standard definition of deference[:]

$$Pr_{\text{Tom}}(P \mid Pr_{\text{Tess}}(P) = v) = v$$

> But why should Tom match his credence in [one] object to Tess's credence in some *other* object?
>
> [(iii)] A further problem—with even deeper repercussions... concerns conditionalization. [Y]our *a priori* connections can change over time, in which case the primary intension of P as uttered by you at one time can be different from the primary intension of P as uttered by you at a different time... conditionalization should not require your credence at t_1 in some object to be constrained by your conditional credence in t_0 in some *other* object. (161-163)

Without providing a full defense of two-dimensionalism from these concerns, I think that what we have covered so far gives grounds for optimism. We can consider Mahtani's concerns in reverse order. For worry (iii)—concerning Conditionalization—I argued above that we *need* diagonalization for a proper version of the norm. For worry (ii)—deference to experts—I have appealed to others who have argued similarly: Gallow, cited above, claims for independent reasons that the proper way to operationalize the Principal Principle, which has exactly the same form as Mahtani's Tom-Tess deference norm, will insert diagonalizing operators into the scope of both probability functions.

Finally, for (i)—attribution of sameness (and difference) of credal content, we can tap Beauty again. Suppose Beauty is a thirder, and it is now Sunday. Consider these two true statements:

(10) Present Beauty's credence that [today is not Tuesday] is 1.

(11) Future Beauty's credence that [today is not Tuesday] is 2/3.

From these two statements, a simple thinking-alike norm of attribution would lead us to infer that Beauty is going to violate conditionalization: there is some P such that her credence in P is going to drop from 1 to something lower. The more complex norm of credal attribution the two-dimensionalist would endorse blocks this inference. But that's a good thing. In the envisioned scenario, Beauty does not update irrationally.

References

Chalmers, David (2011a). "Frege's Puzzle and the Objects of Credence." *Mind* 120(479), 587–635.

Chalmers, David (2011b). "Propositions and Attitude Ascriptions: A Fregean Account." *Noûs* 45. 595–639.

Elga, Adam (2000). "Self-locating belief and the Sleeping Beauty problem." *Analysis* 60. 143–147.

Gallow, J. Dmitri (2023). "Two-Dimensional *de se* Chance Deference." *Australasian Journal of Philosophy* 1–19. doi:10.1080/00048402.2023.2169945.

Horgan, Terry (2002). "The Two-Envelope Paradox, Nonstandard Expected Utility, and the Intensionality of Probability." *Noûs* 34. 578–603.

Kaplan, David (1989). "Demonstratives." In Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes from Kaplan*, Oxford University Press.

Kripke, Saul (1980). *Naming and Necessity*. Harvard University Press.

Lewis, David (1971). "A Subjectivist's Guide to Objective Chance." *Studies in Inductive Logic and Probability 2* .

Stalnaker, Robert (1978). "Assertion." In *Context and Content*, 78–95. Oxford University Press. In Stalnaker (1999).

Stalnaker, Robert (1999). *Context and Content.* Oxford, Oxford University Press.

Melissa Fusco is an Assistant Professor of Philosophy at Columbia University.