Delusions and other question-sensitive beliefs

PLM workshop on *Delusions in Language and Mind*
Hana Kalpak · Stockholm University · October 24 2020

- Clinically and historically, the dominant view of delusions is *doxastic*: delusions are belief-states. According to the *DSM-5* (2013), delusions are "fixed beliefs that are not amenable to change in light of conflicting evidence."
- Critics of the doxastic view (e.g., Berrios, 1991; Campbell, 2001; Currie, 2000) argue that delusions do not behave sufficiently like beliefs to qualify as such: delusions are
  - insufficiently *inferentially integrated* with subjects' other beliefs,
  - insufficiently *responsive to evidence*, and/or
  - insufficiently *action guiding*

  to be belief states.
- Defendants of the doxastic view may respond by denying that belief (essentially) has the functional role assumed by the opponents.
- Arguments to this effect have appealed to independently motivated treatments of belief systems as *fragmented*: as comprised of compartmentalized subsystems which need not be mutually consistent, and which need not be simultaneously accessible to the subject (Cherniak, 1986; Lewis, 1982; Stalnaker, 1984).
- While these types of analyses are typically motivated by non-pathological cognitive phenomena (such as problems of logical omniscience), they also assign a functional role to beliefs that seems to better accommodates the alleged lack of inferential integration, sensitivity to evidence, and action guidance exhibited by delusions.
- This talk investigates the extent to which delusions can be accommodated as beliefs within Yalcin (2016)'s analysis of belief as both fragmented and relativized to *questions*, understood as partitions of logical space.
- It will also be degree an investigation of this framework itself, aiming in particular to take some first steps towards formulating satisfactory constraints on *rationality* of fragmented, question-relative beliefs.
- This is independently desirable: Models of belief systems as fragmented are typically not coupled with any *positive* thesis about belief rationality. Thus such models run the risk of classifying too many beliefs as rational, and too few belief revisions as necessary.
- This problem is highlighted by the question of the doxastic status of delusions. Those who want to classify delusions as beliefs typically do not want to accommodate them as just any type of beliefs, but minimally as *irrational* beliefs (cf. Bortolotti (2010)).

> **What I will suggest**: In a framework where beliefs are both fragmented and relativized to questions, we can define some intuitive rationality constraints on belief using a standard notion of *question entailment*. These constraints are independently motivated, yet serve also to classify delusions as irrational beliefs.

Outline

- Section 1 summarizes the debate regarding the doxastic status of delusions, and the appeal to fragmented models of belief;
- Section 2 introduces the "simple fragmented model" of belief, and notes the obvious limitations of it;
- Section 3 introduces Yalcin (2016)'s model of belief as *question-sensitive*,
- Section 4 proposes two rationality constraints for the question-sensitive model, and relates it back to the debate regarding the doxastic status of delusions;
- Section 5 concludes.

There is also an Appendix summarizing some formal details.

1 Delusions as belief-states

- Bortolotti (2010) identifies and summarizes three types of arguments put forward in support of the thesis that delusions are not beliefs, of which I will discuss two.

> **Argument #1:** Beliefs are integrated with other beliefs: we normally accept the obvious logical implications of our beliefs, and seek to obtain and maintain consistency among our beliefs. If delusions are not integrated with a person's beliefs, then they are not beliefs.

- Some reported delusions are simply so bizarre that it is difficult to see how they could be maintained by any remotely functional person in a consistent belief state.
- Patients suffering from delusions may report that they have a nuclear power station inside of their body (David, 1990), that they are in both Boston and Paris at once (Weinstein and Kahn, 1955), or that they are dead (a characteristic of *Cotard's delusion*).
- Others are not bizarre in content, but contradict other beliefs made explicit by the subject. Breen, Caine, Coltheart, Hendy and Roberts (2000) describe a patient (DB) suffering from reduplicative paramnesia, who claimed (incorrectly) that her husband was a patient in the same hospital as her, yet affirmed (correctly) that her husband was dead.

> **Argument #2:** Beliefs are responsive to evidence: we form beliefs based on our available evidence, and revise extant beliefs in the face of obvious counterevidence. If delusions are not responsive to evidence, then they are not beliefs.

- The 'evidence' that delusional patients mention in support of their delusions are generally insufficient, and often lack obvious connection to the content of the delusion.
- Jordan, Lockert, Johnson-Warren, Cabell, Cooke, Greer and Howe (2006) report of a subject with *erotomania*—the delusion that another person (usually of higher status) is in love with them. She claims as evidence for her delusion that the other person sends her messages proving his love for her in the form of (e.g.) license plates from other U.S. states and the color purple.

- Delusions are also characteristically maintained in the face of what would normally be accepted as incontrovertible counterevidence.
- Brett-Jones, Garety and Hemsley (1987) cite the case of a patient who insisted that his psychologist was deaf-mute, yet was happy to converse with her.
- Young and Leafhead (1996) report of a patient with Cotard's delusion—the delusion of being dead or non-existent—, called "JK":

  > We therefore asked [JK], during the period when she claimed to be dead, whether she could feel her heart beat, whether she could feel hot or cold [...]. She said she could. We suggested that such feelings surely represented evidence that she was not dead, but alive. JK said that since she had such feelings even though she was dead, they clearly did not represent evidence she was alive. (Young and Leafhead, 1996, pp.157-158)

  In sum, each type of argument assumes that beliefs have a certain functional role, and proceeds to argue from the observation that delusions (too often, at least) fail to fulfill this role, to the conclusion that delusions are not beliefs.
- This gives the defendant of a doxastic view of delusion two natural points of defense:
  - denying that delusions *lack* the functional role attributed to beliefs; Strategy 1
  - denying that belief *has* the functional role attributed to beliefs. Strategy 2
- I will focus on arguments following Strategy 2, which question that beliefs are (constitutively) integrated, responsive to evidence, and action guiding.
- In particular, I'm interested in arguments that appeal to belief as *fragmented* or *compartmentalized* in order to establish a more generous functional role for belief.
- For instance, Bayne and Pacherie (2005) suggest that (following (Cherniak, 1986) limitations of our short-term memory force us to store beliefs in relatively independent subsystems, such that we are not generally able to recall and use (for inference or action) beliefs from all subsystems at once.
- Then "failure to draw certain inferences or notice and resolve certain inconsistencies" concerning a purported belief is not counterevidence to its belief-status: inconsistencies between two subsystems may go unnoticed because the systems are not simultaneously active.
- Thus the hypothesis of fragmentation is thought to help against **Argument #1**, by allowing failures of belief-integration.
- It could potentially also be used against part of **Argument #2**: perhaps belief revision does not happen to all "subsystems" at once, but rather revision happens "locally", with respect to the subsystems active at the time where revision is prompted.
- Bortolotti (2010) argues that fragmentation hypotheses (like Bayne and Pacherie's, at least) are insufficient responses to the anti-doxastic arguments.
- Fragmentation in this sense will allow for beliefs to be *diachronically* inconsistent, i.e. to be inconsistent because they are not simultaneously "active" or attended to.

- Delusions however seem to be characteristically involved also in *synchronic* inconsistency, i.e., inconsistency between beliefs that are (near-)simultaneously attended to.
- Making a deluded subject aware of the inconsistency between her delusion and another belief is characteristically insufficient for making her reject the delusion.
- Bortolotti argues that delusions are best understood as *irrational* beliefs, partly because they figure in specifically *synchronic* inconsistencies.
- To accommodate delusions in a fragmented model of belief, then, we want a model that allows
  - both diachronic and synchronic inconsistencies between beliefs, and — Requirement 1
  - diachronic but not synchronic inconsistencies to be classified as *rational*. — Requirement 2
- Bortolotti notes that fragmented models of belief generally fail to meet Requirement 1 (and so also Requirement 2)—typically, such models do not say much about the synchronic case, or at most, they tacitly assume that synchronic consistency is constitutive of belief.
- Later in this talk, I will suggest how Yalcin (2016)'s fragmented model of belief can be extended in order to meet both Requirement 1 and Requirement 2.

2 The simple fragmented model

- Since Hintikka (1962): the representational (informational) content of a state of belief is as a *set of possible worlds*, such that an agent *believes* a proposition p just in case p is true in all of these worlds.
- This view of the representational content of belief is much too simple. For instance, it implies *Closure under conjunction* and *Closure under entailment*, both of which are widely agreed to make for a much too idealized notion of belief.

> **Closure under conjunction.** If a believes p and a believes q, then a believes $p \cap q$.

> **Closure under entailment.** If a believes p and p entails q (i.e., p is true in all q-worlds), then a believes q.

- To deal with issues like these, Stalnaker (1984) suggests a type of *fragmentation* of belief: that our beliefs correspond not to one single set of belief worlds, but to multiple.
- On this picture, the representational content of a state of belief is a *set of sets of possible worlds*, compartmentalized from each other so that one may believe that p with respect to one, but that $\neg p$ with respect to another, and so forth.
- Going forward, I will refer to this picture as the *simple fragmented model* of belief, and follow Yalcin (forthcoming) in calling the set of sets a *doxastic state*, and its members *belief states*.
- Plenty of everyday scenarios invite the simple fragmented model, or something much like it. Lewis (1982, p. 436) offered the following, "based on a true story":

Lewis. David K. Lewis used to think that (i) Nassau Street ran roughly east-west, and that it was roughly parallel to the nearby railroad, and that (ii) the railroad ran roughly north-south, and that it was roughly parallel to Nassau Street. He did not believe in the blatantly inconsistent conjunction of these two: the different beliefs came into action in different situations, and the whole system of beliefs never manifested itself all at once. Once it did, straightaway his beliefs changed: thereonafter, he believed that Nassau Street and the railroad both run roughly northeast-southwest.

- This scenario motivates a fragmented model by showing that it invalidates *Closure under conjunction, fragmented*:

Closure under conjunction, fragmented. If a believes p with respect to some belief state, and a believes q with respect to some belief state, then a believes $p \cap q$ with respect to some belief state.

- With respect to one fragment, Lewis believed p : *Nassau Street runs east-west and is parallel to the railroad*; with respect to another, Lewis believed q : *The railroad runs north-south and is parallel to Nassau Street*. Yet with respect to no fragment does Lewis believe that $p \cap q$ (the inconsistent state).

2.1 Two problems for the simple fragmented model

- **Problem #1.** *Closure under entailment, fragmented* is valid on the model:

Closure under entailment, fragmented. If a believes p with respect to some belief state, and p entails q, then a believes q with respect to some belief state.

- The individual belief states are still sets of possible worlds, and must as such be closed under entailment, given a standard treatment of propositions as sets of worlds and of entailment as set inclusion.

- But *Closure under entailment, fragmented* is still problematic. Stalnaker (1984) illustrates the problem with the following case (the draft discusses an additional case):

The absent-minded detective. The absent-minded detective believes the butler did it, but totally overlooks the possibility it was the chauffeur. Does the detective believe that the chauffeur did not do it? Intuitively, no; or at least there is something off about this characterization of his mental state.

- The simple fragmented model would classify the detective as believing that *The chauffeur did not do it*, since this is entailed by *The butler (and nobody else) did it*.

- **Problem #2:** The simple fragmented model assumes that the content of a belief state is consistent, or at least that this is *rationally required* (Yalcin, forthcoming).

> **Fragmentary coherence (FC).** It is rationally required that the belief states of a doxastic state be internally consistent (i.e., that they be non-empty).

- However, this is hardly sufficient: we also want some *interfragmentary* rationality constraint on beliefs, governing how belief states may co-exist in a rational doxastic state.
- If mutually inconsistent belief states can always co-exist peacefully within one and the same doxastic state, it is mysterious why Lewis does—and intuitively, *ought*—to revise his beliefs when he discovered the inconsistency.
- By the same token, an interfragmentary constraint is required in order classify the deluded patient's refusal or inability to revise inconsistent beliefs as irrational; i.e., in order to classify *synchronic* inconsistency as irrational.
- Yalcin (forthcoming) considers various possible rationality constraints for fragmented models, discussing in particular the principle of *Interfragmentary coherence* (IC).

> **Interfragmentary coherence (IC).** It is rationally required that the belief states of a doxastic state be consistent.

- This principle is (rightly, I think) rejected—it would classify Lewis as irrational.
- I will propose a weaker rationality constraint, defined for Yalcin (2016)'s model of belief as not only fragmented, but question-sensitive.
- This model was proposed in order to avoid **Problem #1** (*Closure under entailment, fragmented*), while preserving a core possible worlds-treatment of belief.
- What we need, he suggests, is not just a model of belief that allows for multiple belief states, but one that also makes a distinction within the beliefs at a given state, according to whether they are *accessible* (roughly, explicit) or *inaccessible* (roughly, implicit) to the subject. To capture this, he makes doxastic states *question-sensitive*.
- I will argue that by making doxastic states question-sensitive, we can also make progress on **Problem #2** of distinguishing rational doxastic states from irrational ones).

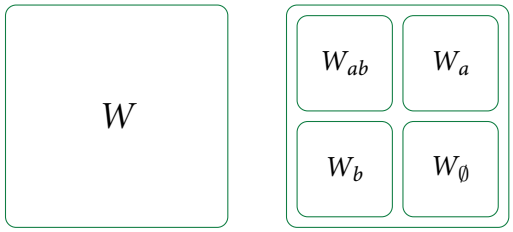
3 The question-sensitive model

- On the *question-sensitive model of belief*, the representational contents of our beliefs are modelled as *doxastic states*: functions from partitions of logical space (*questions*) to subpartitions (*belief states*), which in turn determine a set of *accessible* and a set of *inaccessible* beliefs.
- Let's unpack this. A *partition* of a space of worlds is a division of the space into mutually exclusive and jointly exhaustive regions.

Partition. A *partition* of a set A is a set $\pi(A)$ of mutually disjoint, nonempty subsets of A, such that that $\bigcup \pi(A) = A$.

Illustration. $\pi_{ab}(W) = \{W_{ab}, W_a, W_b, W_\emptyset\}$, illustrated in Figure 1(b), is a partition of the logical space of possible worlds into the sets

- W_{ab} of all and only worlds where **both Ana and Bea** are happy,
- W_a of all and only worlds where **Ana but not Bea** is happy,
- W_b of all and only worlds where **Bea but not Ana** is happy, and
- $W_\emptyset$ of all and only worlds where **neither Ana nor Bea** is happy.



(a) W (b) $\pi_{ab}(W)$

Figure 1: The logical space W and a four-cell partition $\pi_{ab}(W)$ of this space.

- A *subpartition* of a partition $\pi(W)$ is just a subset of $\pi(W)$.
- A doxastic state maps a partition to one of its subpartitions. This is the agent's *belief state* with respect to the partition.

Doxastic state. The *doxastic state* of an agent a, denoted by $\mathcal{B}_a$, is a (partial) function from partitions to belief states.

Belief state. A *belief state* relative to a partition $\pi(W)$ is a subpartition of $\pi(W)$.

- A belief state corresponds to the set of coarse-grained possibilities, out of those distinguished by a certain partition, that an agent has not ruled out.
- The *union* of the belief state of an agent a with respect to a partition $\pi(W)$ is the set of *belief worlds* of a with respect to $\pi(W)$: the set of fine-grained possibilities not ruled out by a, with respect to the distinctions made by $\pi(W)$.

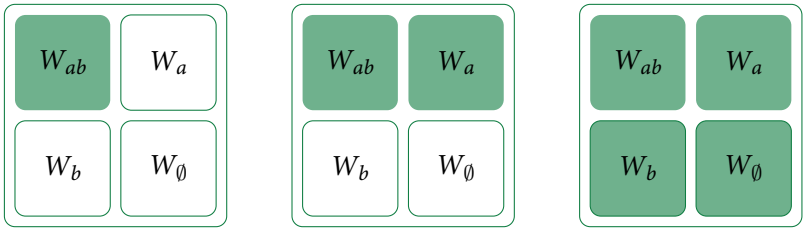
Belief worlds. The set of *belief worlds* of an agent a with respect to a partition $\pi(W)$ is the set $\bigcup \mathcal{B}_a(\pi(W))$.

- The belief worlds determine a set of *beliefs* as usual: Any proposition true across all belief worlds is a belief of the agent in question.

Beliefs. A proposition p is a *belief* at a belief state $\mathcal{B}_a(\pi(W))$ just in case $\bigcup \mathcal{B}_a(\pi(W)) \subseteq p$.

Illustration. The filled cells in Figure 2(a) correspond to the elements of the belief states determined by three different doxastic states.

- Figure 2(a): given a choice between the four indicated possibilities, believes that the actual world is among those where both Ana and Bea are happy.
- Figure 2(b): correspond to the elements of the belief state of someone who, given the same distinction between possibilities, believes that Ana is happy (this is true according to both W_{ab} and W_a), but is unsure as to whether Bea is, too (this is only true according to W_{ab}).
- The filled cells in Figure 2(c) correspond to the contents of the "ignorant" belief state with respect to the distinction at hand: the belief state of someone who has no idea whatsoever about who (out of Ana and Bea) is happy, and thus can exclude no cell as a possible container of actuality.



(a) "Both Ana and Bea are happy."

(b) "Ana is happy, and Bea might be, too."

(c) "No idea."

Figure 2: Three different belief states with respect to $\pi_{ab}(W)$.

- Partitions of W are the formal objects used to represent *question intensions* in the canonical Groenendijk and Stokhof (1984) *partition semantics* for questions.
- In brief, an interrogative sentence Q is taken to denote, at a world w, the proposition corresponding to the *true exhaustive answer* to Q at w.
- For instance, at a world w where Ana but not Bea is happy, the question (1) denotes the proposition W_a mentioned above. This proposition is clearly *true* at w, and *exhaustive* in the sense that it specifies for each individual in the *wh*-words restrictor (viz. Ana, Bea), whether or not the individual is happy.

(1) Who (of Ana and Bea) is happy?

- The intension of (1) is thus a function from worlds to propositions, each of which specifies both whether Ana is happy, and whether Bea is happy.
- There are exactly four such propositions—W_{ab}, W_a, W_b, and $W_\emptyset$—which together cover the logical space. Thus, the intension of (1) is the (characteristic function of the) partition $\pi_{ab}(W)$.

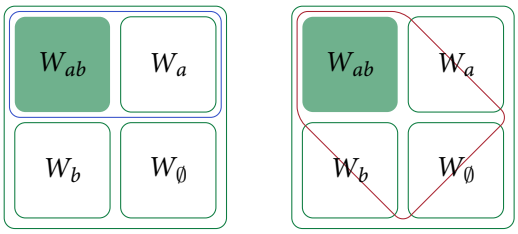
- Thus partitions can be seen as *questions* (like "Who is happy?"), and belief states as (sets of) their *answers* (like "(Only) Ana is happy"), making a doxastic state a *function from questions to (sets of) answers*.[1]
- A proposition is *foregrounded* by a question if it corresponds to the union of elements from the question. Otherwise, it is *backgrounded*.

> **Foregrounded and backgrounded propositions.** A question *foregrounds* all propositions in its union closure, and *backgrounds* all other propositions.

- The *accessible beliefs* of an agent with belief state $\mathcal{B}_a(\pi(W))$ are the propositions that are *true* throughout $\bigcup \mathcal{B}_a(\pi(W))$, and foregrounded by $\pi(W)$.
- The *inaccessible beliefs* are the propositions that are *true* throughout $\bigcup \mathcal{B}_a(\pi(W))$, and backgrounded by $\pi(W)$.

> **Accessible and inaccessible beliefs.** A proposition p is an *accessible belief* at a belief state $\mathcal{B}(\pi(W))$ just in case $\bigcup \mathcal{B}(\pi(W)) \subseteq p$ and p is foregrounded by $\pi(W)$. A proposition p is an *inaccessible belief* at $\mathcal{B}(\pi(W))$ just in case $\bigcup \mathcal{B}(\pi(W)) \subseteq p$ and p is backgrounded by $\pi(W)$.

Illustration. The singleton belief state depicted in Figure 3(a), corresponding to the belief state of someone who believes that *Both Ana and Bea are happy*, makes accessible the belief corresponding to the union of the cells in the blue rectangle, i.e. the belief that *Ana is happy, and Bea might be, too*. In Figure 3(b), the red lines mark the borders of a proposition cutting across the cells of the partition, including all worlds from W_{ab} but only some worlds from the other cells. This proposition is an *inaccessible* belief at the given belief state.



(a) Accessible belief. (b) Inaccessible belief.

Figure 3: Examples of an accessible (blue lines) and an inaccessible (red lines) belief given the partition $\pi_{ab}(W)$ and belief state $\{W_{ab}\}$.

3.1 Closure principles of questions-sensitive beliefs

- Like the simple fragmented model, the question-sensitive model of belief invalidates *Closure under conjunction, fragmented*: one may believe that p wrt. one question $\pi(W)$ and that p' wrt. another question $\pi(W)$, without believing that $p \cap p'$ with respect to any question.

[1] This is *not* to say that the partitions in the domains of our doxastic states are always easy/possible to express by interrogatives in natural language. It is also not to say that interrogatives in natural language necessarily denote partitions—they, as well as the "questions" in the domains of our doxastic state, could be modelled differently.

> **Closure under conjunction, fragmented.** If a believes p with respect to some belief state, and a believes q with respect to some belief state, then a believes $p \cap q$ with respect to some belief state.

- Also like the simple fragmented model, the question-sensitive model of belief *validates* the *Closure under necessary entailment, fragmented*. It is still impossible for a doxastic state to map a partition to a belief state where some proposition p is true, but some proposition q entailed by p is not (this is immediate since then $p \subseteq q$).

> **Closure under necessary entailment, fragmented.** If a believes p with respect to some belief state, and p entails q, then a believes q with respect to some belief state.

- However, it invalidates a weaker thesis, which I will call *Closure under accessible necessary entailment* (CANE):[2]

> **Closure under accessible necessary entailment.** If a accessibly believes p with respect to some belief state, and p entails q, then a accessibly believes q with respect to some belief state.

- This is "some kind of progress, speaking to at least part of what troubles us about closure under entailment" (Yalcin, 2016).
- In particular, it gets at what's troubling us about the case of the absent-minded detective: this is a counterexample to (CANE).

> **The absent-minded detective.** The absent-minded detective believes that the butler did it, but totally overlooks the possibility it was the chauffeur. Does the detective believe that the chauffeur did not do it? Intuitively, no; or at least there is something off about this characterization of his mental state.

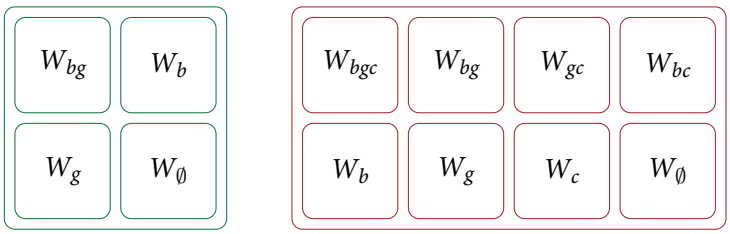
- The absent-minded detective has arrived to his conclusion by considering the wrong question—perhaps something like (2), illustrated in Figure 4(a)—and might come to a different conclusion by considering a question like (3), illustrated in 4(b).

(2) Who (among the butler and the gardener) did it?

(3) Who (among the butler, the gardener, and the chauffeur) did it?

- The detective accessibly believes with respect to (2) that W_b : *the butler (and nobody else) did it*, which of course entails that the chauffeur did not do it. Thus the accessible belief in W_b implies a belief in $W_{\neg c}$: *the chauffeur did not do it*. But $W_{\neg c}$ is not an *accessible* belief: $W_{\neg c}$ is not in the union closure of (2). Thus, this is a counterexample to (CANE).
- So, while the question-sensitive believer is bound to believe any logical consequences of her beliefs, she is not bound to believe them *accessibly*. Thus the model of belief as question-sensitive takes us forward with respect to **Problem #1** discussed above.

[2] (CANE) corresponds to Yalcin (2016)'s (E^4).



(a) *Who (among the butler and the gardener) did it?*

(b) *Who (among the butler, the gardener, and the chauffeur) did it?*

Figure 4: Two questions relevant to the detective.

- Next, I will suggest how the question-sensitive model could also take us forward with respect to **Problem #2**, concerning rationality of fragmented doxastic states.

4 Constraining rationality

- The rationality constraint I will propose will be formulated using the notion of *question entailment*.
- The question (expressed by) (4-a) entails (the one expressed by) (4-b), in the sense that by fully answering (4-a), we also fully answer (4-b).

(4) a. What's Ana's favorite color?
Answers: Ana's favorite color is {green/blue/red/yellow/...}
b. Is Ana's favorite color blue?
Answers: Ana's favorite color is blue, Ana's favorite color is not blue

- For instance, by establishing that *Ana's favorite color is green*, we establish that *Ana's favorite color is not blue*, viz. the negative answer to (4-b).
- On Groenendijk and Stokhof (1984)'s partition semantics for questions, *question entailment* is defined as the relation $\models$.

> **Question entailment.** For two questions φ, ψ : $\varphi \models \psi$ iff for all interpretations I : $[\![\varphi]\!]_I \subseteq [\![\psi]\!]_I$

That is, a question φ entails a question ψ iff, for any interpretation I, the (true exhaustive) answer to φ at I entails the (true exhaustive) answer to ψ at I. (Recall that the *extension* of a question is its *true exhaustive answer* at the world of evaluation.)

Note: The *conjunction* of two questions will always entail its conjuncts. For instance, the question in (5) is the conjunction of the questions (6), (7).

(5) Is Ana's favorite color blue, and is Bea's favorite color yellow?

(6) Is Ana's favorite color blue?

(7) Is Bea's favorite color yellow?

- Intuitively, any full answer to the conjunctive question (e.g., *Ana's favorite color is blue, but Bea's favorite color isn't yellow*) is a full (overinformative) answer to the individual polar questions. Thus, (5) entails both (6) and (7).
- On the partition semantics for questions, this follows from a definition of *question conjunction* ($\wedge$) as *pairwise intersection* (see Appendix).
- We can now use the notion of question entailment to define rationality constraints on question-sensitive belief states, that are weaker than the constraint (IC) considered and rejected by Yalcin (forthcoming).
- I propose, minimally, the following two principles:

> **Interfragmentary inclusion under entailment (IIE).** If a doxastic state $\mathcal{B}$ is defined for two questions φ, ψ such that $\varphi \vDash \psi$, then it is rationally required that $\mathcal{B}(\varphi) \subseteq \mathcal{B}(\psi)$.

- That is: It is rationally required that the belief states with respect to two questions related by entailment are related by inclusion, in the same direction.

> **Synchronic closure under conjunction (SCC).** If a doxastic state $\mathcal{B}_a$ is defined for two questions φ, ψ and a synchronically attends to both φ and ψ, then it is rationally required that $\mathcal{B}_a$ is defined for $\varphi \wedge \psi$.

- That is: It is rationally required that the domain of a doxastic state be closed under conjunction of questions synchronically attended to. (I know, this is really underspecified, but you get the broad idea.)
- I will also assume a version of *Fragmentary coherence* (FC):

> **Fragmentary coherence (FC).** It is rationally required that the belief states of a doxastic state be internally consistent (i.e., that they be non-empty).

Illustration. According to (FC) + (IIE) + (SCC), if I believe

- that **Ana is happy**, w.r.t. the question *Is Ana happy?*, and
- that **Bea is happy** w.r.t. the question *Is Bea happy?*,

then if I attend to the two questions simultaneously, it is rationally required that I also believe

- that **Both (Ana and Bea) are happy** w.r.t. the question *Who (of Ana and Bea) is happy?* (or equivalently, *Is Ana happy, and is Bea happy?*).

4.1 (IIE), (SCC) and Problem #2

- Together with the constraint (FC), (IIE) and (SCC) will allow rational *diachronic* inconsistencies, but disallow rational *synchronic* inconsistencies.
- Let's look first at what this says about the Lewis-case.

Lewis. David K. Lewis used to think that (i) Nassau Street ran roughly east-west, and that it was roughly parallel to the nearby railroad, and that (ii) the railroad ran roughly north-south, and that it was roughly parallel to Nassau Street. He did not believe in the blatantly inconsistent conjunction of these two: the different beliefs came into action in different situations, and the whole system of beliefs never manifested itself all at once. Once it did, straightaway his beliefs changed: thereonafter, he believed that Nassau Street and the railroad both run roughly northeast-southwest.

- Recall that **Problem #2** with respect to the Lewis-case was that the fragmented model did nothing to explain *why* Lewis did (and *rightly* so) revise his beliefs once he discovered an inconsistency.
- In terms of the model at hand, Lewis believed with respect to the question q_1 of *the orientation of Nassau Street and its position relative to the railroad*, that

  (8) Nassau Street runs east-west and is roughly parallell to the railroad.

  and with respect to the question q_2 of *the orientation of the railroad and its position relative to Nassau Street*, that

  (9) The railroad runs north-south and is roughly parallell to Nassau Street.

- His doxastic state $\mathcal{B}_l$ was not necessarily irrational: q_1 and q_2 are not related by entailment in the sense defined.
- However, once he attends to both questions at once (for some reason or other), it is rational to *update* $\mathcal{B}_l$ into a state defined for $q_1 \wedge q_2$, given (SCC).
- Now, a doxastic state which is just like $\mathcal{B}_l$, only that its domain includes $q_1 \wedge q_2$, *would* be irrational, given (IIE) + (FC).
- Such a state (call it $\mathcal{B}_{l'}$) must map $q_1 \wedge q_2$ to some non-empty belief state, in order not to violate (FC).
- But in this case, $\mathcal{B}_{l'}$ is irrational according to (IIE): no non-empty state can be a subset of both the belief state $\mathcal{B}_{l'}(q_1) = \{[\![(8)]\!]\}$ and the belief state $\mathcal{B}_{l'}(q_2) = \{[\![(9)]\!]\}$, since $[\![(8)]\!]$ and $[\![(9)]\!]$ are disjoint. **Upshot:** $\mathcal{B}_{l'}$ is irrational.
- With the proposed rationality constraints, the question-sensitive model captures the dynamics of Lewis' belief change as a **rational update** of the doxastic state $\mathcal{B}_l$ with the question $q_1 \wedge q_2$, in the sense that the updated doxastic state conforms to (FC) + (IIE) + (SCC).
- Instead of updating to some $\mathcal{B}_{l'}$-type state, Lewis updates into a state $\mathcal{B}_{l*}$ where his beliefs with respect to q_1 and q_2 are *revised*, in order to allow for a rational belief with respect to their conjunction: $\mathcal{B}_{l*}(q_1 \wedge q_2) = (10)$.

  (10) Both Nassau Street and the railroad run roughly northeast-southwest.

- This distinguishes the Lewis-case from cases where one or more beliefs involved in the inconsistency is delusional.

- To exemplify, I will consider **the reduplicative paramnesiac** DB (Breen et al., 2000), who claimed over the course of one and the same interview, both that her husband was an alive patient in the same hospital as her, and that he was dead (which was correct).
- Her doxastic state, call it $\mathcal{B}_{db}$, can be modelled as defined on two polar (yes/no) questions:

  (11) Is my husband dead? (= *Is my husband alive?*)
  a. My husband is dead.

  (12) Is my husband alive and a patient in the same hospital as me?
  a. My husband is alive and a patient in the same hospital as me.

- With respect to (11), DB in $\mathcal{B}_{db}$ believes the answer (11-a), and with respect to (12), she believes the answer (12-a). Obviously, these beliefs are inconsistent.
- Assume that she "attends" to these questions synchronically.
- Then if $\mathcal{B}_{bd}$ is **not** defined on (11) $\wedge$ (12), then this state is irrational by (SCC).
- I will say that in this case, DB is *externally synchronically inconsistent*: she fails to at all update her doxastic state with the conjunction of two questions synchronically attended to.
- If $\mathcal{B}_{bd}$ **is** defined on (11) $\wedge$ (12), then this state is irrational by (IIE) + (FC). This state is analogous to $\mathcal{B}_{l'}$ considered above: $\mathcal{B}_{bd}$((11) $\wedge$ (12)) will be empty (inconsistent) if obeying (IIE), since it entails both (11) and (12).
- I will say that in the former case, DB is *internally synchronically inconsistent*: she succeeds in updating her doxastic state with the conjunction of two questions simultaneously (almost, at least) attended to. However, she fails to perform a *rational* update of her doxastic state.
- Thus, we can distinguish two types of synchronic inconsistency in the model: external and internal. Both are classified as irrational.

5 Conclusion

- By adding the principles (IIE) + (ACC) to (FC), the question-sensitive model satisfies both Requirement 1 and Requirement 2, which stated that a model of delusions-as-irrational-beliefs must
  - allow both diachronic and synchronic inconsistencies between beliefs, and Requirement 1
  - allow diachronic but not synchronic inconsistencies to be classified as *rational*. Requirement 2

  Two notes: *First*, also certain diachronic inconsistencies will be classified as irrational, in virtue of (IIE) + (FC) alone.
- The case at hand illustrates this: DB's *diachronic* inconsistency would also be irrational. (12) itself entails (11), meaning that DB's beliefs with respect to the former should rationally entail her beliefs with respect tot the latter.

- Another type of intuitively irrational inconsistency, involving failure of closure under "conceptual entailments", is discussed in the draft.
- *Second*, some intuitively irrational cases of *belief revision* exhibited by delusional subjects are *not* captured by the model, as it stands.
- Recall JK, the Cotard patient who refused to revise the (purported) belief that she was dead, and instead chose to deny the connection between having a heart beat and being dead.
- The doxastic state resulting from refusing to revise an obviously implausible belief, at the expense of an obviously plausible one, is not irrational by the standards suggested here.
- Capturing the irrationality of this move would require additional constraints on rational belief revision, likely in a probabilistic version of the framework, which I leave for future work.

References

Bayne, T. and Pacherie, E.: 2005, In defence of the doxastic conception of delusions, *Mind & Language* **20**(2), 163–188.

Berrios, G. E.: 1991, Delusions as "wrong beliefs": A conceptual history, *The British Journal of Psychiatry* **159**(S14), 6–13.

Bortolotti, L.: 2010, *Delusions and other irrational beliefs*, Oxford University Press.

Breen, N., Caine, D., Coltheart, M., Hendy, J. and Roberts, C.: 2000, Towards an understanding of delusions of misidentification: Four case studies, *Mind & Language* **15**(1), 74–110.

Brett-Jones, J., Garety, P. and Hemsley, D.: 1987, Measuring delusional experiences: A method and its application, *British Journal of Clinical Psychology* **26**(4), 257–265.

Campbell, J.: 2001, Rationality, meaning, and the analysis of delusion, *Philosophy, Psychiatry, & Psychology* **8**(2), 89–100.

Cherniak, C.: 1986, *Minimal rationality*, Mit Press.

Currie, G.: 2000, Imagination, delusion and hallucinations, *Mind & language* **15**(1), 168–183.

David, A. S.: 1990, Insight and psychosis, *The British Journal of Psychiatry* **156**(6), 798–808.

DSM-5: 2013, Arlington, Va.

Groenendijk, J. and Stokhof, M.: 1984, *On the semantics of questions and the pragmatics of answers*, PhD thesis, University of Amsterdam.

Groenendijk, J. and Stokhof, M.: 1997, Questions, *in* J. van Benthem and A. ter Meulen (eds), *Handbook of Logic and Language*.

Hintikka, J.: 1962, Knowledge and belief: an introduction to the logic of the two notions.

Jordan, H. W., Lockert, E. W., Johnson-Warren, M., Cabell, C., Cooke, T., Greer, W. and Howe, G.: 2006, Erotomania revisited: thirty-four years later., *Journal of the National Medical Association* **98**(5), 787.

Lewis, D.: 1982, Logic for equivocators, *Noûs* **16**(3), 431–441.

Stalnaker, R.: 1984, Inquiry.

Weinstein, E. A. and Kahn, R. L.: 1955, Denial of illness: Symbolic and physiological aspects.

Yalcin, S.: 2016, Belief as question-sensitive, *Philosophy and Phenomenological Research* **97**(1), 23–47.

Yalcin, S.: forthcoming, Fragmented but rational, *in* C. Borgoni, D. Kindermann and A. Onofri (eds), *The Fragmented Mind*, Oxford University Press.

Young, A. W. and Leafhead, K. M.: 1996, Betwixt life and death: Case studies of the cotard delusion, *Method in madness: Case studies in cognitive neuropsychiatry* pp. 147–171.

A Question semantics basics

Following Groenendijk and Stokhof (1997), we represent interrogative sentences as first-order sentences '$?\vec{x}.\varphi$'. '?' is a question operator binding the length n variable sequence '$\vec{x}$', where $n =$ the number of *wh*-phrases in the interrogative. Thus $n = 2$ for a question like *Who likes what?*; $n = 1$ for a question like *Who is happy?*, and $n = 0$ for a polar question like *Is Ana happy?*. Then the *extension* of a question '$?\vec{x}.\varphi$' is defined as in (13-a), where $g'[\vec{x}]g$ means that g' differs from g at most in the values assigned to $\vec{x}$. This expresses that '$?\vec{x}.\varphi$' has the same extension in two worlds w, w' just in case the exact same value assignments to $\vec{x}$ make φ true in w' and w.

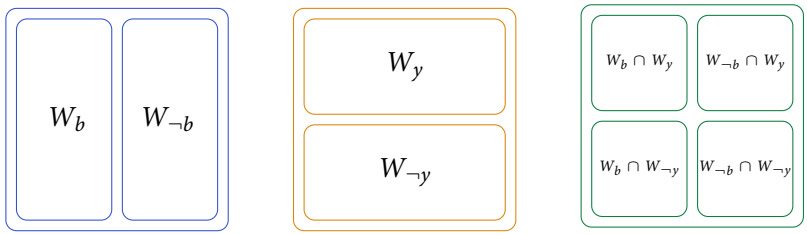
(13) a. $[\![?\vec{x}.\varphi]\!]_{M,g,w} = \{w' \mid \forall g'[\vec{x}]g : [\![\varphi]\!]_{M,w,g'} = [\![\varphi]\!]_{M,w',g'}\}$
b. $[\![?\vec{x}.\varphi]\!]_{M,g} = \{[\![?\vec{x}.\varphi]\!]_{M,g,w} \mid w \in W\}$

The *intension* of '$?\vec{x}.\varphi$' then corresponds to (13-b): a partition of W into sets of worlds that agree on which assignments of values to $\vec{x}$ make φ true.

The conjunction of two questions is defined as *pairwise intersection*:

Question conjunction. For two questions φ, ψ :
$$[\![\varphi \wedge \psi]\!] = \{p \cap p' \mid p \in \varphi \text{ and } p' \in \psi\}.$$

Figure 5(c) depicts the result of conjoining the two questions *Is Ana's favorite color blue?* and *Is Bea's favorite color green?*.



(a) $\pi_b(W)$: *Is Ana's favorite color blue?*

(b) $\pi_y(W)$: *Is Bea's favorite color yellow?*

(c) $\pi_b(W) \wedge \pi_y(W)$

Figure 5: Two polar questions and their conjunction.

It should also be noted (this is relevant if we want to model the rationality of the *absent-minded detective, for instance*) that if φ, ψ are two *wh*-questions such that the *domain* of ψ—intuitively, the set of individuals it asks about—is a subset of φ, then φ entails ψ.

For instance, a full answer to (15) (e.g., that *Ana and Cem are happy, but not Bea*) entails a full answer to (14) (i.e., the answer that *Ana but not Bea is happy*). This follows from the definition of a question.

(14) Who (of Ana and Bea) are happy?

(15) Who (of Ana, Bea, and Cem) are happy?