The intuitive invalidity of the pain-in-mouth argument

MICHELLE LIU (D)

1. Introduction

The pain-in-mouth argument (Block 1983: 517) presents a puzzle about pain. Consider the following inference:

There is a pain in my fingertip.

The fingertip is in my mouth.

Therefore there is a pain in my mouth.

Intuitively, the argument is invalid, but philosophers disagree over what precisely explains the intuitive invalidity at issue, and furthermore they often take their proposals to support different philosophical theories of pain.

Tye (1995, 2002, 2005) contends that the argument is invalid because 'pain' creates an intensional context. He takes his proposal to support representationalism about pain, according to which pain experiences represent tissue damage. Noordhof (2001, 2002, 2005), agreeing with Block (1983), thinks that the 'in' of 'pain in X' is non-spatial and the phrase should be understood as describing X as being in a particular state.

In a recent paper (2019), appealing to empirical results, Reuter, Sienhold and Sytsma ('RSS' henceforth) argue against these two existing proposals and put forward an *implicature account*. They contend that the conclusion in the argument carries the conversational implicature that *there is something wrong with the speaker's mouth*. But the premisses don't carry this implicature. This explains why we have the intuition that the argument fails, though the argument, on this proposal, is strictly speaking valid. RRS (2019: 81) take their proposal to support the bodily view of pain, according to which pains are states of the body, not states of the mind (see Reuter and Sytsma 2020).

This paper argues against the implicature account. It offers two arguments – one theoretical and one empirical – against the proposal (§2). Drawing on further empirical evidence, it shows that pain reports using locative locutions, such as 'There is a pain in my mouth', are intuitively understood by ordinary English speakers as *entailing* corresponding predicative locutions, such as 'My mouth hurts'. The paper thus vindicates a rather simple and unsurprising solution to the pain-in-mouth puzzle: the *entailment account*, according to which the argument seems invalid because the conclusion is understood as entailing something that cannot be inferred from the premisses

(§3). The philosophical implications of this proposed solution are also drawn (§4).

2. Against the implicature account

RSS (2019: 74) model the pain-in-mouth argument on the following kinds of arguments:

There is tissue damage in my finger.

The finger is in my mouth.

Therefore, there is tissue damage in my mouth.

There is an inflammation in my finger.

The finger is in my mouth.

Therefore, there is an inflammation in my mouth.

Like the pain-in-mouth argument, the above arguments also seem invalid. RSS's (2019) explanation for the intuitive invalidity of these arguments crucially lies with their claim that the following utterances, 'in most conversational settings', carry the conversational implicature that *there is something wrong with the speaker's mouth*:

There is a pain in my mouth.

There is tissue damage in my mouth.

There is an inflammation in my mouth.

According to RSS (2019: 74), we are inclined to judge that the above arguments fail because their conclusions all carry the *generalized conversational implicature* 'that something is wrong with the speaker's mouth, while the premisses carry no such implicature'. However, RSS's proposal fails to provide a full explanation for the intuitive invalidity of these arguments.

A generalized conversational implicature is carried by an utterance of a proposition in most ordinary contexts of utterance (Blome-Tillmann 2013: 178). Consider the following two well-known examples of generalized conversational implicatures from Grice 1975: 56:

- (a) 'I went to a college yesterday.'
- (b) 'I'm meeting a woman for dinner tonight.'

In most ordinary contexts, (a) carries the implicature that the college I went to is not mine; (b) carries the implicature that the woman is not my wife. Now consider the following arguments:

(A1)

I went somewhere yesterday.

This somewhere is a college.

I went to a college yesterday.

(A2)

Someone is meeting a woman for dinner tonight.

This someone is I.

I am meeting a woman for dinner tonight.

In both arguments, it seems that the conclusion carries a generalized conversational implicature that is not carried by the premisses. Regarding (A1), we can easily imagine ordinary conversational contexts where a subject. upon being asked whether she was home vesterday, replies naturally with 'I went somewhere yesterday', which does not implicate that the place the speaker went to was not her college. Similarly, regarding (A2), there seem to be ordinary contexts where the utterance of 'Someone is meeting a woman for dinner tonight' does not carry the conversational implicature generally associated with the conclusion. Such a context could be one where the gender of the person (in this case, woman) or the purpose of the meeting (for dinner) or the time of the meeting (tonight) is contextually salient. Regarding the first option, for instance, upon seeing a lone woman sitting at a table for two in a restaurant typically full of men, one waiter says to another: 'It looks like someone is meeting a woman for dinner tonight', which does not implicate that the woman is not this someone's wife.

RSS explain the intuitive invalidity of the pain-in-mouth argument in terms of a generalized conversational implicature carried by the conclusion but not carried by the premisses. Arguments (A1) and (A2) are like the pain-in-mouth argument in this respect. However, we do not judge them to be invalid nor do we find the conclusions misleading. More generally, just because the conclusion of an argument carries a certain generalized conversational implicature that is not carried by the premisses, it does not follow that we find the argument intuitively invalid. So, RSS's account, as it stands, cannot adequately explain the intuitive invalidity of the pain-in-mouth argument.

Furthermore RSS's claim that utterances such as 'There is a pain / tissue damage / an inflammation in my mouth' carry the conversational implicature that 'There is something wrong with my mouth' is questionable. Conversational implicatures are usually thought of as explicitly cancellable (Grice 1975: 44, Blome-Tillmann 2008, 2013). If utterance 'P' merely conversationally implicates Q, then the utterance 'P, (but) not Q' is admissible or at least not outright contradictory. This is known as 'the cancellability test'. Consider the following example from Blome-Tillmann (2013: 170)

A: Are you going to the party tonight?

B: I don't like parties.

B's utterance carries the conversational implicature that B won't go to the party tonight. But this implicature is explicitly cancellable. B would not be contradicting herself if she uttered the following:

(1) I don't like parties, but I'll go to this one anyway.

Generalized conversational implicatures like 'I'm meeting a woman for dinner tonight' also pass the cancellability test. The following utterance is admissible:

(2) I'm meeting a woman for dinner tonight. It's my wife.

The cancellability test helps to distinguish conversational implicatures from semantic entailments. Semantic entailments fail to cancel (Blome-Tillmann 2013: 172):

- (3) #Elliot is a bachelor (i.e. a single man), but he is married.
- (4) #The general killed himself, but he was not dead.

Given the standard cancellability test, and given RRS's implicature account, namely that utterances like 'There is a pain / tissue damage / an inflammation in my mouth' carry the conversational implicature that 'There is something wrong with my mouth', the following sentences should be admissible:

- (5) There is a pain in my mouth, but there is nothing wrong with my mouth.
- (6) There is tissue damage in my mouth, but there is nothing wrong with my mouth.
- (7) There is an inflammation in my mouth, but there is nothing wrong with my mouth.

Two observations are to be made here. First, (6) and (7) do not sound admissible at first pass, or at least do not sound as admissible as (1) and (2), which are cancellable conversational implicatures. Second, (5) may or may not sound inadmissible depending on whether one is aware of cases of referred pains, where pains are perceived in locations that are not the source of the painful stimuli.¹

To test these two observations, experimental data were collected from 121 participants on Amazon Mechanical Turk. Participants were first given the following vignette:

According to RSS's study, participants showed an increase in approval of the conclusions of the pain-in-mouth argument, and corresponding arguments featuring 'tissue damage' and 'inflammation', upon being prompted to ignore the misleadingness of relevant conclusions and rate whether they 'technically speaking' follow from the premisses. RSS take this to show that the relevant implicature is cancellable to some extent. However, the increase is small (see 2019: 80–81). For instance, the average rating for the acceptability of the conclusion in the pain-in-mouth argument after the prompt still falls short of the midpoint (increased from *M*=2.10 to *M*=3.79). (RSS used a seven-point Likert scale, where 1 is 'Strongly Disagree', 7 is 'Strongly Agree', and 4 is 'Neither Agree nor Disagree'.) One might wonder whether the increase is simply due to other factors, for example demand characteristics, rather than the cancellability of the implicature.

In this task, you will be asked to judge whether some sentences make sense or not. You are asked to pay attention to the content of the sentence, i.e. what the sentence says.

In order to understand the task, please read the following example:

Consider the sentence: 'The tabletop is rectangular, but it has three sides'. The sentence doesn't make sense at all. It expresses a contradiction, because something cannot be rectangular while having three sides at the same time.

Participants were then asked to rate whether sentences (1)–(7) make sense.² The sentences were randomly ordered. A seven-point Likert scale was used where 1 means 'The sentence makes perfect sense' and 7 means 'The sentence makes no sense at all.'

Results were compared across three groups: *cancellable conversational implicatures*, (1) and (2); *contradictions*, (3) and (4); and *bodily conditions*, (5)–(7). Most participants gave low ratings to (1) (M=1.76, SD=1.29) and (2) (M=1.43, SD=0.97). Most participants gave high ratings to the two instances of contradictions, (3) (M=6.75, SD=0.99) and (4) (M=6.85, SD=0.72), as well as to the three cases of bodily conditions, (5) (M=5.45, SD=1.85), (6) (M=6.19, SD=1.48) and (7) (M=6.13, SD=1.45).

Results showed that participants did not treat the three sentences about bodily conditions like cancellable conversational implicatures.³ The results are presented as Figure 1 below.

Results also showed that the average rating of (5) was lower than those of (6) and (7) ((5) M = 5.45, SD = 1.85; (6) M = 6.19, SD = 1.48 and (7) M = 6.13, SD = 1.45). A Participants were asked to specify their reasons for

- 2 Participants were 62.5% women and 100% native speakers of English with an average formal education of 15.6 years.
- 3 The ratings for each of the three groups were averaged first. A pairwise comparison showed significant differences between the average ratings of bodily conditions and cancellable conversational implicatures (t(87) = 25.28, p < 0.001). The comparison between the average ratings for bodily conditions versus contradictions had a smaller effect size (Cohen's d = 0.86), whereas the comparison between the average ratings for bodily conditions versus cancellable conversational implicatures had a larger effect size (Cohen's d = 3.87).

To ensure statistical differences did not arise due to the different number of items in each category, results were also compared between the average ratings of the two sentences in the cancellable conversational implicatures category (1 and 2) and subsets of two sentences in the bodily conditions category. Three pairwise comparisons showed significant differences between the average ratings of cancellable conversational implicatures (1 and 2) and the average ratings of bodily conditions (5 and 6: t(87) = 23.23, p < 0.001; 5 and 7: t(87) = 22.20, p < 0.001; 6 and 7: t(87) = 28.03, p < 0.001).

4 A pairwise comparison showed significant differences between the ratings of (5) and (6) (t(87) = 3.73, p < 0.001), and between the ratings of (5) and (7) (t(87) = 4.26, p < 0.001).

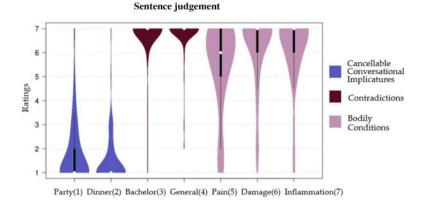


Figure 1. A violin plot displaying the distribution shapes of ratings of whether sentences (1)–(7) make sense. White dots indicate median values. Thick black bars represent interquartile ranges. Thin black lines extended from black bars represent the upper and lower adjacent values in the data, where the upper adjacent value is the largest observation that is less than or equal to the third quartile plus 1.5 times the interquartile range, and the lower adjacent value is the smallest observation that is greater than or equal to the first quartile minus 1.5 times the interquartile range.

their ratings of (5), and were given three options: (a) 'I thought if someone has a pain in his/her mouth, then there must be something wrong with the person's mouth'; (b) 'I am aware of cases of referred pains, where one can have a pain in one part of the body but there is nothing wrong with that body part'; (c) 'Other'. Further examination of the participants' ratings revealed that the lower average rating for (5) was driven by participants whose reasoning aligned with choice (b). Specifically, participants who chose (b) had an average rating of 3.94 (SD = 1.91), whereas participants who chose (a) and (c) had average ratings of 6.39 (SD = 1.00) and 6.67 (SD = 0.58) respectively. The breakdown of the participants who chose (a), (b) and (c) was 58.0%, 38.6%% and 3.4% respectively.

Although (6) and (7) received high average ratings, results did not show that participants treated them like semantic contradictions such as (3) and (4).⁵ A plausible explanation, as noted in one participant's feedback, is that the utterance 'There is something wrong with my mouth' might indicate a level of severity. A subject can have a pain / an inflammation / tissue damage in the mouth, while insisting that there is nothing wrong with her mouth, meaning that there is nothing serious or nothing to worry about regarding her mouth. It could also

A pairwise comparison showed no significant difference between the ratings of (6) and (7) (t(87) = 0.44, p = 0.66).

⁵ A pairwise comparison did show significant differences between average ratings of (6) and (7) on the one hand and contradictions on the other (t(87) = 4.66, p < 0.001).

be that not everyone associates pains, tissue damage or inflammations with negative valence.

But that (5)–(7) were not treated like typical cases of semantic contradictions is not evidence that they involve conversational implicatures. On the contrary, the empirical data suggest the opposite. Given RSS's implicature account, one would naturally expect (5)–(7) to be judged as similar to (1) and (2), rather than (3) and (4). However, this was not the case. RSS might insist that conversational implicatures are cancellable to different extents and some are very hard to cancel. But this is not an adequate response. The burden is on RSS to explain why (5)–(7), supposing they involve conversational implicatures, are so hard to cancel. In other words, RSS must explain why these are special such that they were judged to be nothing like standard cases of cancellable conversational implicatures but more like contradictions.

3. The entailment account

What then explains the intuitive failure of the pain-in-mouth argument? In their recent paper, Liu and Klein (2020) draw attention to the distinction between two distinct kinds of pain report in English: (i) the locative locution, for example 'There is a pain in my back', which at the level of surface grammar describes pains as things located in body parts; and (ii) the predicative locution, for example 'My back hurts', whose surface grammar attributes a state to a body part. A number of philosophers have previously suggested that the two kinds of pain report are closely related (Hyman 2003, Bain 2007). A plausible explanation for the pain-in-mouth puzzle is the following: in English, pain reports using locative locutions are intuitively understood as entailing corresponding predicative locutions - the conclusion of the painin-mouth argument 'There is a pain in my mouth' entails that 'My mouth hurts'. This consequence should also be entailed by the premisses. But it is not. So the conclusion of the argument does not follow from the premisses, and the pain-in-mouth argument is intuitively judged to be invalid. Call this 'the entailment account'.

To empirically test this hypothesis, the same participants were asked to rate, again on the scale of 1–7, whether the following two sentences make sense:

- (8) There is a pain in my mouth, but my mouth doesn't hurt.
- (9) My mouth hurts, but there is no pain in my mouth.

Participants were then explicitly asked whether the following two sentences mean the same thing:

6 Liu and Klein (2020) note four features of the locative locution: 'pain' takes the prepositional phrase 'in NP'; is countable; permits an existential construction and also a possessive construction.

- (10) There is a pain in my mouth.
- (11) My mouth hurts.

Results showed that the average ratings for (8) and (9) were 6.76 (SD = 0.86) and 6.67 (SD = 0.81) respectively, and that participants treated (8) and (9) just like the two cases of semantic contradictions, (3) and (4). The results are summarized in Figure 2.

For the two sentences (10) and (11), 80.7% responded that they mean the same thing.⁸ The results showed that while (10) and (11) are thought of as mutually entailing, they are not universally treated as having the same meaning.

A second experiment was also done with a new set of participants. Data were collected from 124 participants on Amazon Mechanical Turk. This time, participants were given six sentences: (3), (4), (8), (9) and two sentences that are clearly non-contradictory. The order of the sentences was randomized. Instead of being asked whether or not the sentences *make sense*, participants were asked whether the sentences are *contradictions* on a scale of 1–7 where 1 means 'This sentence is definitely not a contradiction' and 7 means 'This sentence is definitely a contradiction.' Similar results were found: (3) (M=6.74, SD=1.10), (4) (M=6.88, SD=0.59), (8) (M=6.80, SD=0.69) and (9) (M=6.74, SD=0.74). Again, participants treated (8) and (9) just like contradictions.¹⁰

If (8) is treated as a contradiction, that is, (10) is thought of as entailing (11), then it is no surprise that the pain-in-mouth argument seems invalid. The underlying reasoning of our judgement regarding the intuitive invalidity of the argument may be spelt out in the following way: if the premisses of the argument entail the conclusion, which is (10), and (10) entails (11), then the

- ANOVA yielded no significant difference among (3), (4), (8) and (9) (F(3, 348) = 0.68, p = 0.58). The two one-sided sample tests (TOST) procedure (Lakens 2017) showed statistical equivalence between the average ratings for contradictions and ratings for (8) ($\Delta_L = -0.20$, $\Delta_U = 0.28$, t(174) = -82.44, p < 0.001), as well as between the average ratings for contradictions and ratings for (9) ($\Delta_L = -0.10$, $\Delta_U = 0.36$, t(174) = -84.19, p < 0.001). (For all equivalence tests conducted for this paper, the lower t-value, associated with the higher p-value, is reported.)
- 8 11.4% thought they do not mean the same, of which 40% gave a rating of 7 to both (8) and (9); 5.7% were not sure; 2.3% gave their own responses.
- 9 The two non-contradictory sentences were: 'If today is Sunday, then yesterday was Saturday'; 'Tom's bookshelf is full of books.' The average ratings for these two sentences were 1.15 (SD = 0.87) and 1.05 (SD = 0.31) respectively.
- 10 The two one-sided sample tests (TOST) procedure showed statistical equivalence between the average ratings for contradictions and ratings for (8) ($\Delta_L = -0.18$, $\Delta_U = 0.20$, t(218) = -102.87, p < 0.001), as well as between the average ratings for contradictions and ratings for (9) ($\Delta_L = -0.13$, $\Delta_U = 0.27$, t(218) = -98.88, p < 0.001). Participants were also asked whether (10) and (11) mean the same thing. This time, 86.3% responded affirmatively.

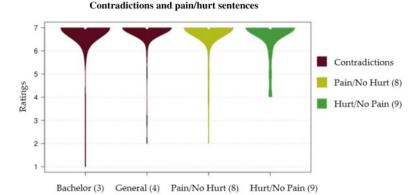


Figure 2. A violin plot displaying the distribution shapes of rating of whether sentences (3), (4), (8) and (9) make sense. White dots indicate median values. For each category, the interquartile range, the distance between the upper and lower quartiles, is 0, and the upper and lower adjacent values are all at 7.

premisses should also entail (11). But they don't, so the premisses don't entail the conclusion and the argument is invalid.

The entailment account proposed here, although similar to some existing accounts in the literature, such as Noordhof's account and the 'paraphrase account' mentioned by Hyman (2003) and Bain (2007), needs to be distinguished from them. According to the latter two proposals, (10) ought to be understood as (11). For Noordhof (2001, 2002), this is so because the relevant 'in' is non-spatial but used in a state-attributing sense. For Hyman (2003: 16–17) and Bain (2007: 182), it is because (10) can be paraphrased as (11) without any loss of meaning. The entailment account is less committal. It does not claim that the 'in' of 'pain in X' is non-spatial. Nor is it committed to the idea that (10) and (11) have the same meaning, however meaning is conceived. It only claims that ordinary English speakers treat (10) as entailing (11), which is supported by empirical data.

Philosophers have compared the pain-in-mouth argument to a number of other arguments that also seem invalid. Notably Noordhof (2001) has compared it with 'the hole-in-box argument' – there is a hole in my shoe; the shoe is in the box; therefore, there is a hole in the box. Tye (1995) has compared it to invalid arguments involving propositional attitude verbs: Tom believes that he is in Vancouver; Vancouver is in Canada; therefore Tom believes that he is in Canada. RSS have compared it, as we have seen, to arguments

11 The claim that the 'in' of 'pain in X' is non-spatial but state-attributing is problematic (see also Hyman 2003, Reuter et al. 2019). There are clear examples of state-attributing uses of 'in' in English: *in doubt*, *in love* etc. In these cases, 'in' is followed by a noun phrase which indicates a state. In the case of 'in the fingertip/mouth', the relevant noun phrases do not indicate states – fingertips and mouths are not states.

involving tissue damage and inflammation. The entailment account provides an explanation as to why the pain-in-mouth argument seems intuitively invalid to us. An advocate of the account is not required to explain the intuitive failures of these other arguments. It could well be the case that different kinds of explanations should be given to account for the intuitive failures in these other arguments. ¹²

4. Conclusion

As we have seen in this paper, the implicature account fails to explain the intuitive failure of the pain-in-mouth argument. In contrast, the entailment account offers a plausible and empirically backed explanation. The argument seems invalid because the conclusion is naturally taken to entail that the speaker's mouth hurts, which cannot be inferred from the premisses. As it stands, this account makes no claims about where *pains* are located and is neutral between the mental and bodily conception of pain.

The entailment account is, however, in tension with the alleged possibility that one can have a pain in a body part without that body part *hurting*. In a number of places (e.g. Sytsma and Reuter 2017, Reuter and Sytsma 2020), Reuter and Sytsma have presented empirical evidence to show that ordinary English speakers are open to the possibility of unfelt pains or pains that don't hurt. The evidence presented in this paper, in contrast, indicates that ordinary English speakers treat the utterance 'There is a pain in my mouth, but my mouth doesn't hurt' as a contradiction, which suggests that they take pains to necessarily hurt.

This tension between the two sets of evidence may plausibly reflect a tension in the (English) folk conception of pain. The overall situation seems to go some way in support of the recently proposed view of Borg et al. that the (English) folk conception of pain is *polyeidic*, that is, 'containing a number of different strands or elements', where 'in different contexts different elements of the concept could be activated, enhanced or supressed' (2020: 30–31). It may well be the case that what drives our intuitive judgement that the pain-in-mouth argument is invalid is the thought that if there is a pain in X then X must hurt. But such a conception of pain may be suppressed in some contexts. It is possible that when considering hypothetical scenarios concerning a particular imaginary subject (who is e.g., distracted, on pain killers etc.) and being explicitly asked in the locative forms whether

¹² With some of these arguments, it may be conjectured that locutions such as 'There is an inflammation in my mouth', 'There is a hole in the box' etc. are also intuitively thought of as entailing propositions that can be expressed in some predicative form: 'My mouth is inflamed', 'My shoe is perforated' etc. which would explain why relevant arguments are also intuitively invalid. But this claim requires separate empirical testing.

¹³ See Borg et al. 2020 for a review of the relevant experimental literature.

there is a pain in the subject's body part or whether the subject has a pain (see Reuter and Sytsma 2020), participants might be triggered to focus on bodily damage and treat pains as concrete physical entities located in body parts, which in turn suppresses their conception of pains as necessarily hurt or felt. However, further study is required to understand the underlying factors that influence our judgements. 14

Funding

This work was supported by the Oxford Philosophical Fellowship Fund Postdoctoral Award.

> University of Hertfordshire, Hatfield AL10 9EU, UK v.liu43@herts.ac.uk

References

Bain, D. 2007. The location of pains. Philosophical Papers 36: 171-205.

Block, N. 1983. Mental pictures and cognitive science. Philosophical Review 92: 499-541.

Blome-Tillmann, M. 2008. Conversational implicature and the cancellability test. Analysis 68: 156-60.

Blome-Tillmann, M. 2013. Conversational implicatures (and how to spot them). Philosophy Compass 8: 170-85.

Borg, E., R. Harrison, J. Stazicker and T. Salomons. 2020. Is the folk concept of pain polydeidic? Mind and Language 35: 29-47.

Grice, H.P. 1975. Logic and conversation. In Syntax and Semantics III: Speech Acts, ed. P. Cole and J.L. Morgan, 41-58. New York: Academic Press.

Hyman, J. 2003. Pains and places. Philosophy 78: 5-24.

Lakens, D. 2017. Equivalence tests: a practical primer for t tests, correlations, and metaanalyses. Social Psychological and Personality Science 8: 355-62.

Liu, M. and C. Klein. 2020. Pain and spatial inclusion: evidence from Mandarin. Analysis 80: 262-72.

Noordhof, P. 2001. In pain. Analysis 61: 95-97.

Noordhof, P. 2002. More in pain. Analysis 62: 153-54.

Noordhof, P. 2005. In a state of pain. In Pain: New Essays on the Nature of Pain and the Methodology of Its Study, ed. M. Aydede, 151-62. Cambridge, MA: MIT Press.

14 I would like to thank an anonymous reviewer for valuable comments. Thanks also to the audience at the Pain, Pleasure and the Method of Cases conference in Bern where the paper was presented. I am also indebted to the following people for helpful discussions and feedback: David Bain, Martin Davies, Domi Dessaix, Jessica Keiser, Luke King-Salter, Michaela Jirout Košová, Luis Ospina-Forero and Kevin Reuter. Special thanks to Jiahe Zhang for advice regarding the experiments.

Reuter, K. and J. Sytsma. 2020. Unfelt pains. Synthese 197: 1777-1801.

Reuter, K., M. Sienhold and J. Sytsma. 2019. Putting pain in its proper place. *Analysis* 79: 72-82.

Sytsma, J. and K. Reuter. 2017. Experimental philosophy of pain. *Journal of Indian Council of Philosophical Research* 34: 611–28.

Tye, M. 1995. A representational theory of pains and their phenomenal character. Philosophical Perspectives 9: 223–39.

Tye, M. 2002. On the location of a pain. Analysis 62: 150-53.

Tye, M. 2005. Another look at representationalism about pain. *In Pain: New Essays on the Nature of Pain and the Methodology of Its Study*, ed. M. Aydede, 99–120. Cambridge, MA: MIT Press.

(Un)knowability and knowledge iteration

SEBASTIAN LIII

The KK principle, in its original form, states that if a subject S knows some proposition p, then S knows that S knows that p. Schematically, where 'K' is the knowledge operator and 'Kp' denotes 'S knows that p', we can formulate the KK principle as follows:

(KK)
$$Kp \supset KKp$$
.

The impressive list of KK proponents in the history of philosophy includes Plato, Aristotle, Aquinas, Spinoza, Locke, Schopenhauer and Hintikka. More recently, however, a number of putative counterexamples have led many philosophers to reconsider their commitments to this iteration principle. Nevertheless some have maintained that, while we should concede KK in its original form, there are versions of the principle that are defensible. The most prominent version of the revised KK principle states that if *S* knows that *p*, then *S* is *in a position* to know that *S* knows that *p*. Since knowing entails being in a position to know but being in a position to know does not entail knowing, this formulation is weaker than KK. Where '*Pp*' denotes '*S* is in a position to know that *p*', we can schematize this version of knowledge iteration as follows:

(PK)
$$Kp \supset PKp$$
.

Principles akin to PK have received endorsements to varying degrees from, for example, McHugh (2010: 231), Greco (2014: 173–74), Stalnaker (2015: 28), Das and Salow (2018: 8) and Goodman and Salow (2018: 184).

1 A more comprehensive list of KK defenders can be found in Hintikka 1962: chapter 5.