

Multiple Review in *Mind & Language*

Author reply

*Draft – please cite published version*

## ***Representation in Cognitive Science – Replies***

**Nicholas Shea**

### **Abstract**

In their constructive reviews, Frances Egan, Randy Gallistel and Steven Gross have raised some important problems for the account of content advanced by Nicholas Shea in *Representation in Cognitive Science* (2018, OUP). Here the author addresses their main challenges as follows. Egan argues that the account includes an unrecognised pragmatic element; and that it makes contents explanatorily otiose. Gallistel raises questions about homomorphism and correlational information. Gross puts the account to work to resolve the dispute about probabilistic contents in perception, but argues that a question remains about whether probabilities are found in the content or instead in the manner of representation.

### **Introduction**

I'm very grateful to all three reviewers for tackling the book so constructively and posing a series of important questions for me to grapple with. I don't have space to mention, let alone answer, all the points raised, and I have a lot to think about still. In my reply I'll try to address the points that are most challenging. These challenges are as follows.

Frances Egan diagnoses a pragmatic element in my accounts of content determination, based as they are partly on causal explanation. Secondly, she argues that the content-fixing properties I advert to can explain behaviour without having to appeal to representational contents. Randy Gallistel contributes a welcome perspective from psychology. Amongst much agreement about the basics, I note a potential difference in the way we rely on homomorphisms. I also explain why my accounts can make do with a very spare form of correlation (probability raising), while recognizing that measures of mutual information and efficient coding have an important role to play in a wider explanatory project. Steven Gross puts my account to work, applying it to the debate about whether probabilities are represented in perception. He also gets down into the details of my view and raises a penetrating question about the distinction between representing probabilistic contents and representing categorical contents in a probabilistic manner.

## (1) Egan: Explanation and Pragmatic Content

The representational theory of mind (RTM) is committed to representations having physical vehicles which interact in virtue of non-semantic properties. Theories of content then face the problem that local factors look set to capture everything that goes on with a system. Frances Egan has a carefully worked-out account of content that grapples with this issue (Egan 2010, 2014). She concludes that computational properties, local to the system, are objective, but the representational content carried by computational vehicles is merely a pragmatic gloss. Examining my account, she diagnoses a pragmatic element at work there too.

My account invites this reaction, since it makes a central appeal to explanation. Content is fixed by exploitable correlations and structural correspondences<sup>1</sup> that figure in unmediated causal explanations. I say that causal explanation is ubiquitous in the sciences. My account therefore shows that representation is no more interest-relative, and hence no less naturalistic, than the properties found in any other science. However, even if causal explanation is an objective dependency relation, Egan argues that there are too many such objective relations that could be relied on in fixing content. My account chooses between them and, Egan argues, it necessarily does so in an interest-relative way. By privileging mediated over unmediated causes of stabilisation I smuggle in pragmatic considerations. That distinction lies in the eye of the beholder.

My response is that the way content is fixed in my account depends only on causal relations between organisms, their outputs, conditions in the environment and stabilising processes. My account specifies which causal relations are involved in fixing content in a non-interest-relative way.

Let me illustrate by reference to the rat navigation case. The rat is at location L in a maze when it observes the location of some food. It goes through a series of internal computations (as described in §5.2) and then sets off in a certain direction, taking an efficient route to the food. How did this complex computational and behavioural disposition get stabilised? Presumably natural selection is part of the story, but so is learning. Let's suppose the rat set off from L in the past and got food from another location. This feedback stabilised (reinforced) the whole system of computation-plus-behaviour.

On that occasion place cell R was firing, and entered into the internal processing which eventually led the rat to set off in a certain direction. Like any vehicle, R will carry a whole host of correlational information. Let's suppose R's tokening correlates with the following conditions:

- C1 The rat is at location L.
- C2 The rat is at a location where there is a little black dot on the floor.
- C3 The rat is at a location where visible features  $V_1$  to  $V_n$  are arranged with angular separations of  $A_1$  to  $A_{n-1}$ .

---

<sup>1</sup> For simplicity, to start with I'll focus just on exploitable correlations.

It is because the rat is at L that setting off in the direction it does gets it efficiently to food. So C1 is a cause of the whole behavioural-cum-computational process getting stabilised. C2 and C3 are not. C2 could be mentioned in a causal explanation of stabilisation, but only if that explanation were supplemented with the information that C2 makes C1 likely; the same for C3. Only C1 features in an unmediated explanation of stabilisation.

Notice that this reasoning is not about how internal states come to correlate with conditions in the environment. The test is about which conditions play a causal role in stabilisation. The 'unmediated' in unmediated correlational information concerns the causal explanation of stabilisation being unmediated, not the correlation.

Opting for conditions that figure causally in stabilisation does make those explanations less opaque. But content is determined, not by the opacity or transparency of the explanation, but by the causal facts. Compare: a longer text tends to be harder to explain, but the length of a text is determined by its word count, not by how difficult it is to explain.

We can put these considerations to work in the case against which previous teleosemantic theories have been so extensively tested, the frog's tongue dart reflex. We can suppose that the case is supplemented, if necessary, to meet my more stringent conditions on content determination. The case generates firm intuitions, which is unhelpful as I think intuitions about content have almost no probative value in this context. Furthermore, even in subpersonal cases, such an early signal is usually used for several purposes, which would push in the direction of contents like *little black thing* rather than *fly*. But as the case is described philosophically, there is a direct throughput connection between circumstances, representation, behavioural output and stabilisation. This invites us to think all at once about how the whole system works, especially what the behaviour is causally responsive to. What we should first be thinking about is how it gets stabilised.

When we focus on stabilisation, the causal story is reasonably clear: the frog darts out its tongue to a particular location, there is a fly there, the frog ingests it, and that whole behavioural disposition is stabilised as a result. This is just a biological-causal story. Various conditions are causes of the behaviour leading to stabilisation (here: leading to survival and reproduction). That includes there being a fly at a certain location. It's being black is not a cause of survival. An explanation of survival could mention that there was a little black thing at the relevant location, but that is only explanatory relative to the background information that most little black things in this environment are flies. An unmediated causal explanation of stabilisation just mentions the causes. Whether a condition is a cause is not observer relative. Granted, there is some indeterminacy in the notion. Maybe it is unclear whether it is the biological category *fly* or the ecological category *flying nutritious object* that figures as a cause. But on no view does an object's being black cause the frog to survive. Compare: for everyday objects volume may correlate with mass, but it is mass rather than volume that causes the weighting scales to tip.

This is the most straightforward kind of case. An internal state  $R$  correlates directly with a condition  $C$  that figures in a causal account of stabilisation and/or robustness. Most cases in cognitive science are more complex. A series of representations  $R_1$  to  $R_n$  are calculated over on the way to producing an internal state  $R$  that correlates with a condition  $C$  that is unmediatedly causally involved in stabilisation. A theory of content needs to account for representations  $R_1$  to  $R_n$  as well.

For example, in the case in §4.7, one vehicle  $R_1$  correlates with the chromatic reflectance properties of surfaces (shorthand: colour), another  $R_2$  with local motion (registered achromatically). Information from an array of these vehicles is combined to produce a vehicle  $R$  that correlates with the direction of motion of an object. In the case in question this was the condition  $C$  which features causally in stabilising the behaviour in question (through learning). The organism's behavioural disposition to reach to such-and-such location was reinforced because there was an object moving in such-and-such direction. What of the conditions concerning colour and local motion? They do not appear in the causal explanation of stabilisation at all. They show up in a different place: in a causal explanation of how it is that  $R$  correlates with condition  $C$ . Tokening of vehicle  $R$  is caused by tokening of vehicles of types  $R_1$  and  $R_2$  in such a way that, given that  $R_1$  correlates with local surface colour and  $R_2$  correlates with local motion,  $R$  correlates with the direction of motion of a whole object (condition  $C$ ).

According to my account all these internal states carry UE information. How so? (In particular: watch out that a pragmatic element isn't being smuggled in.)  $R$  raises the probability of  $C$  –  $P(C|R) > P(C)$  – because  $R$  is produced by an internal algorithm over  $R_1$  to  $R_n$  which marches in step with a set of corresponding external conditions  $C_1$  to  $C_n$  which together raise the probability of  $C$ . More carefully: there is an internal algorithm by which tokening of  $R_1$  to  $R_n$  in a certain sequence with a certain timing causes  $R$  to be tokened; each raises the probability of a corresponding condition  $C_1$  to  $C_n$  such that the obtaining of conditions  $C_1$  to  $C_n$  in a corresponding sequence and timing is the causal basis for the probabilistic connection between  $R$  and  $C$  (i.e. for  $P(C|R)$ ). This complex collection of causal facts underpins an unmediated causal explanation of the system's performance of its task functions. That is what it is for exploitable correlational information carried by the states  $R_1$  to  $R_n$  to figure, through implementing an algorithm, in an unmediated explanation of the system's performance of task functions (i.e. to meet the definition of UE information).

Thus, a condition  $C$  qualifies as UE information, hence as the content of a representation, in virtue of a constellation of causal facts, none of which are in the eye of the explainer. This complex web is relevant to content because it gives us a causal explanation of task function performance. For ease of exposition, UE information is defined terms of unmediated causal explanation, through implementing an algorithm, of task function performance. However, that is just convenient shorthand for the obtaining of the complex of causal facts spelt out in the previous paragraph.

This gives us a recipe for content determination for prior states  $R_1$  to  $R_n$ . Amongst all the very many conditions with which they respectively correlate we need to find

conditions  $C_1$  to  $C_n$  such that the obtaining of those conditions together (in a certain sequence, with a certain timing) makes  $C$  more likely, and does so in a way that marches in step with the way  $R_1$  to  $R_n$  cause  $R$  to be tokened, thereby underpinning the correlation between  $R$  and  $C$ . That is a demanding constraint. And it is a matter of meeting a complex causal condition. Considerations of explanatory interests don't enter into it – at least not unless facts about causal relations in biology are observer-dependent in general. In the latter case, naturalism is still secured because my account would make representational content no more pragmatic than the properties in found in other special sciences.

Egan's second objection is that content attribution is not essential to the explanation of an organism's success [m/s p. 13]. We can instead appeal to non-semantic facts about robust correlations between internal components and states of affairs distal to the organism. However, as Egan notes, robust correlations are just *some* of the facts that make up my cluster [p. 14]. That is not enough to home in on the natural cluster which I argue underpins the abundance of representation-using systems and the corresponding utility of representational explanation (§3.2). It is a more specific cluster of properties that gives content its explanatory bite. Accordingly, robust correlations do not take us all the way to explaining successful behaviour. They can sometimes explain how an organism manages to produce a certain outcome robustly. They don't tell us, however, why some robustly-produced outcomes rather than others should count as successes (§6.4a).

Egan argues that there are too many robust correlations between internal components and distal facts. So determined, content is unavoidably indeterminate. I agree, but what narrows down candidate contents for me is the rest of the cluster: the connection to stabilisation, and the need to find correlations for a collection of components that march in step with a computation which serves to produce stabilised outcomes. Egan is right that we need something more than robust correlations, but where she appeals to the pragmatic interests of those who seek to explain behaviour, I point to a collection of further causal facts. If representational contents are fixed in this more complex way, as I claim, then they are not explanatorily otiose.

## **(2) Gallistel: Homomorphism and Information**

Since the task of giving an empirically well-motivated theory of content is clearly an interdisciplinary endeavour, it is a privilege to have the benefit of Randy Gallistel's incisive commentary; the more so as the discussion of representation in his landmark book, *The Organization of Learning* (1990), partly inspired the account I advance in chapter 5. We are clearly in broad agreement that content (meaning) depends in part on how a representation is processed in the internal computational machinery: 'The manipulations performed on the symbols must make sense given [their] referents' [m/s p. 1]. This contrasts with theories of content that seek to found content on symbol-world relations independently of how the symbols are processed.

Gallistel appeals here to homomorphism: ‘the way in which [a symbol’s] form interacts with the physical form of the computational machinery must be such that the results of a computation are homomorphic to the properties of the world to which the symbols operated on may be assigned referents ...’ [p. 9]. The computational process mimics a corresponding real-world process (Gallistel & King 2010, p. 250). It is important to distinguish this from another place where homomorphism can figure in a theory of content, that is in giving rise to a structural representation. The former is a matter of the computations that occur between a set of representations. It is about the box and arrow diagram which shows how computational steps are carried out on vehicles (hence how information flows through the system). The latter is a matter of the structures on which computations take place. It applies to a broadly map-like structure where the relations between the elements are themselves vehicles, that is, they affect how downstream computations unfold. My account of structural correspondence relies on homomorphisms of the latter kind (UE structural correspondence).

Most of Gallistel’s commentary concerns the former kind, an elaboration of the just-mentioned idea that the computations performed on representations must make sense in the light of their content. On the other hand, when he contrasts spatial representation with the colour representation system, he highlights a structural feature in the latter sense. Relations between representations in the colour space are not used computationally, whereas relations between representations of spatial locations are so used [m/s p. 10]. I argue that it is only when a relation between representations is used computationally that it is a candidate to be part of a structural representation (to figure in a UE structural correspondence). Gallistel agrees (e.g. Gallistel & King 2010, p. 164, last para.). When the relations between items in a data structure are relied on computationally for the way they correspond to relations in the world (e.g. relations of sweetness: Gallistel & King 2010, p. 165) we have a case of structural representation. I have argued that we should not confuse these two ways in which something called a ‘structure’ is relevant to content: ‘computational structure’ is not a case of structural representation (pp. 138-9). Computational structure presupposes regularities in the world, but only implicitly (Shea 2015), not in a way that can be used by other downstream computations. Gallistel uses ‘functioning isomorphism’ indiscriminately for both cases (Gallistel 1990, Gallistel & King 2010) but we should separate them sharply because their respective roles in fixing content are quite different.

Moving on to his most trenchant criticism, Gallistel objects to my reliance on the very weak notion of correlational information, which only requires that there is a world state whose probability conditional on tokening an internal vehicle is different from its unconditional probability. He suggests that I need to build more onto this notion, and argues that, if I do, I will run into considerable difficulties with specifying the strength of a correlation [m/s p. 2]. There are problems with formulating correlation strength in terms of a correlation coefficient, flowing from the well-known difficulties in formalizing the notion of contingency that is found in animal learning theory (Gibbon et al. 1974, Gallistel et al. 2014). Measuring contingency in an animal learning

experiment depends on counting the frequency of non-CS trials, but it is hard to define how often the conditioned stimulus was *not* presented.

My response is that I don't want or need a stronger notion of correlation. My definition of correlational information does not fall prey to the difficulties that arise with defining contingency. It does not turn on what happens when a vehicle R is not tokened. We just need the unconditional probability that some condition C obtains and the conditional probability that it obtains given that R is tokened.<sup>2</sup> There is still the question of the relevant reference class, but that is a separate issue (which I deal with, see p. 79). This notion of correlational information is indeed weak, but the way to overcome that is not to find some stronger notion of correlation which will deliver content, but to turn to further factors. One factor is a consideration that Gallistel and I agree is relevant to meaning, namely the constraint that contents must march in step with the physically-realised transitions that occur over their vehicles. A further factor in my account – a very significant restriction – is the need for a link to task functions.

Gallistel suggests I turn to mutual information. Mutual information is also very liberal, and faces the further problem that it is defined over a set of representations and a corresponding set of conditions in the world. If we take one item from that set, say a particular pattern of neural firing, and want to know which state or states of the world constitute its content, the mutual information carried by the set of possible neural states from which it derives does not tell us in any straightforward way. A natural move is to consider how much that particular pattern of neural firing affects the probability of possible world states  $C_i$  (Skyrms 2010; Shea, Godfrey-Smith & Cao 2018).<sup>3</sup> That is certainly a useful and relevant measure, but it inherits all the indeterminacy of correlational information.

Mutual information is an important tool for a different project, that of working out why information processing in the brain has been configured a particular way; and correlatively for predicting how information processing is likely to be configured to solve a given problem. It is central to my account that representations are created to achieve task functions. So I wholeheartedly agree with Gallistel that the brain creates meaning when it creates representations ('sets of possible messages') and configures computational operations on them.<sup>4</sup> It is not a passive receiver of messages from the world [cf. p. 5]. I also agree that it is interesting to ask why the brain creates one solution rather than another, including to determine how much information the brain obtains about a source, and why it preserves only selected information (as in Gallistel's colour example).

However, it is not the task of a theory of content to explain why a representational system is configured a certain way: why one set of possible representations is generated rather than another. Accounts of content need to tell us what a system represents, given the way it is configured. But the question of why the representations and computations are configured as they are is a good one. Here the Shannon-inspired

---

<sup>2</sup> Note that the value of  $P(C|R)$  does not depend on or entail any particular value of  $P(C|\neg R)$ .

<sup>3</sup> I.e. the values of  $P(C_i|R) / P(C_i)$ .

<sup>4</sup> As modelled by artificial neural networks, (Shea 2007).

signalling framework is helpful (Skyrms 2010, Godfrey-Smith 2012, 2014, Shea et al. 2018), and considerations of efficient coding are often enlightening (Martínez, forthcoming). Epistemically, considering what a system is good at coding may help us to work out what it is representing. However, the efficient coding hypothesis is not a good basis for the metaphysics of content determination. What is represented depends on how a system is actually configured, which depends on a range of constraints, and need not be the most efficient solution.

### (3) Gross: Probabilistic Representation

Anyone would want to see their theory put to use, so I am delighted that Steven Gross can rely on it to resolve the dispute about probabilistic contents in perception (Gross & Flombaum 2017, Block 2018). He argues that my theory overcomes the ‘mere sensitivity’ challenge and implies that perceptual representation is probabilistic. However, he goes on to argue, that conclusion is consistent with two different kinds of probabilistic representation. The probabilities could be located in the content, so that a state has a content like  $P(C) = x$ , for a worldly condition  $C$ , or alternatively in the mode or manner with which a non-probabilistic content is represented, something like a degree of belief that  $C$  obtains. This is a subtle and interesting question.<sup>5</sup> I won’t be able to do it justice here. I’m largely in agreement with what Gross says, so I’ll simply raise a few further points.

With conceptual representations the issue turns on whether the thinker deploys probabilistic concepts. With non-conceptual representations it is harder to see what the distinction consists in. Representations are getting their content through a combination of their relations to things in the world and their functional-computational role. Perhaps an account that computes with probabilistic contents can be recast as a computation over categorical contents represented in a probabilistic manner. So a first possibility is that, for non-conceptual representations, the manner-content distinction cannot be drawn with respect to probabilities; or that the issue indeterminate when asked of real systems.

My account does not in fact fix content by considering which content-assignments explain behaviour (e.g. considering which representational explanation works best). Instead, we look at how and why behavioural outputs were stabilised, and consider the role of exploitable correlations with external conditions, and the role of those conditions in causing robust outcomes and stabilisation. So probabilistic contents arise when a vehicle carries fine-grained exploitable correlational information, and its tokening induces a probability distribution over a range of world states which unmediatedly explains how and why a pattern of behaviour was stabilised.

For example, suppose tokening of  $R$  is caused by objects moving in the centre of the visual field. Across the occasions when a vehicle of type  $R$  was tokened during stabilisation, the object encountered was moving at various speeds, falling into a

---

<sup>5</sup> See Moss (2018) for an extended argument in favour of probabilistic contents instead of degrees of belief at the personal level.



normal distribution with mean 100 degrees per second and standard deviation of 10 degrees per second. Suppose that this interacted with the probabilistic distributions corresponding to other internal states  $R'$ ,  $R''$ , etc. to generate actions that were rewarded in suitably-related environmental conditions. Then  $R$  ends up getting a probabilistic content: its content is a normal distribution of object speeds with mean 100 and standard deviation 10 degrees per second.

Those kinds of considerations underpin my argument that probabilistic contents arise in one of the case studies (§4.8). Gross is right to press the question whether that is automatic – whether probabilities will always end up figuring in the *content*. One central issue here is whether the degree of match between represented and actual probabilities plausibly plays a role in explaining success and failure of behaviour. In my example the system was stabilised partly because  $R$  set up a certain probability distribution over object speeds. Suppose now that something about the environment has changed so that the actual distribution of speeds when  $R$  is tokened is different. Perhaps the object is now on average moving at 95 degrees per second, with a standard deviation of 15 degrees per second. The resulting behaviour will now go wrong a proportion of the time, depending on how  $R$  is computed with. The distance between the represented distribution and the actual distribution should explain how often the behaviour goes wrong, or by how much. One test of the putative probabilistic contents is that they should figure in that kind of explanation.<sup>6</sup>

It is not entirely clear to me how cases of probabilistic manner of representation will arise, if that is supposed to contrast with cases of probabilistic contents. I agree with Gross that none of the four considerations he discusses for drawing the contrast is decisive. A small preliminary disagreement first. Gross argues that, when it comes to manner of representation, there is no distinction to be made between mere sensitivity and manner, since sensitivity is a matter of functional role and manner of representation is determined by functional role [m/s p. 7 (end of §6)]. However, manner of representation is not a matter of total functional role. A representation can have functional features that neither contribute to its content nor to its manner. Cases of priming can fall into that category. Furthermore, information carried by the timing of a representation's tokening can affect downstream processing without having computational significance. A fortiori, mere sensitivity to a probabilistic feature (carrying what I called fine-grained exploitable correlations) need not determine the manner with which a content is represented. A representational vehicle can have sensitivities with no downstream consequences, or with downstream consequences that are irrelevant to its computational role hence to its manner of representation.

Nevertheless, we are thinking about cases where probabilistic information has a computationally-significant effect and asking, in respect of those, whether the probabilities show up in the content or the manner. Even if my account entails probabilistic representation in some cases (contents like  $P(C)=x$ ), we should ask if that

---

<sup>6</sup> We can also use a distance measure to define the degree of correctness *on an occasion*, with respect to a particular object moving at a certain speed. For the KL divergence (§4.8) with a discrete distribution this will just depend on the probability assigned by the content to the world state that obtains. Other measures are possible (Shea, Godfrey-Smith & Cao 2018).

is the right result – if it draws the distinction the right way. I agree that representational sophistication is unlikely to settle the issue when we are considering non-conceptual representations. Nor will the issue turn straightforwardly on whether there really are objective probabilities in the world. If there are none, the subjective probability for the representing system, of world states given the tokening of a representation R, is still a good candidate to figure in an explanation of stabilisation. If there are both objective and subjective probabilities, an obvious move would be to put objective probabilities in the content and restrict subjective probabilities to the manner, but I doubt that the issue can be disposed of that simply.

My discussion of manner of representation in chapter 7 seems to suggest that a necessary condition for treating manner separately from content is for the same representational vehicle (same content, on the narrow conception) be reused in more than one functional role, with different consequences. For example, at the level of conceptual representation, the same mental sentence can be used to express an occurrent belief or an occurrent desire, with the two operating in reasoning quite differently. I designed a distinction between descriptive and directive content that doesn't require such reuse. So a system can have directive contents about a condition C that it is not capable of representing descriptively. However, where there is reuse, its existence does not imply that the difference has to show up as manner rather than content. Where the same vehicle can be combined with multiple probability markers, the probability marker is an equally good candidate to be a vehicle of an aspect of content. I agree with Gross that, were different aspects of a representation to be marked with different probabilities, that would argue in favour of a content interpretation. For example, two features, colour and motion say, may be bound to the same perceived object, but with different probability estimates associated with each feature.<sup>7</sup> Then we should think of the probabilities in the content.

The reason I gave for my terminological choice in chapter 7 may not amount to a good argument that vehicle reuse is required if a functional difference is to count as manner rather than content. Or it may not apply to the case of probabilities. Even if it does, we will be faced with many cases where there is reuse, with one aspect of a vehicle reflecting probabilities. Since reuse allows but does not entail that the difference is one of manner, for these cases the question remains open whether the representation has a probabilistic content or whether it represents a categorical content in a probabilistic manner.

I hope we may move towards a resolution by looking closely at whether there is a probabilistic content whose degree of accuracy explains behaviour. There does not seem to be any requirement in a calculus of credences, updated by Bayesian principles, that the subjective probabilities assigned at each stage align with

---

<sup>7</sup> Stocker & Simoncelli (2008) analysed an experiment in which participants had to judge the overall direction of motion of an array of moving dots. They found evidence that different probabilities were represented for different portions of the array (some of which were thrown away following an interim decision as to whether the motion was to the left or right of a reference line). If the representation of the array counts as a single representation, then this is evidence that different aspects of it are represented with different probabilistic contents.

probabilities in the world, even where they exist. On the other hand, if the probabilities that figure in Bayesian computations are a matter of the agent's model of representation-world relations, then the degree of match between the posterior probabilities on which the agent bases a decision and the state of the world may help explain behavioural success / failure, even if a further norm applies, that of updating in accordance with Bayesian conditionalization. Furthermore, the degree of match may be explanatorily significant even if on the worldly side of the relation there is just a categorical matter of fact (effectively, a degenerate probability distribution).<sup>8</sup> So one way of resolving our question could be that all and only such cases count as having probabilistic contents. Where the match between the world state and a probability distribution attached to a representation systematically explains success and failure of behaviour we have probabilistic contents; otherwise, probabilistic manner. If that proves to be tenable, it would furnish a test that is applicable to non-conceptual representations.

There may also be some mileage to the distinction in a system that marks both uncertainty and risk. There could be two probabilistic aspects attached to a reusable vehicle, one with a functional role connected to epistemic actions and the other with a functional role restricted to instrumental calculations, e.g. deciding what to bet on. We might then find that the system becomes more certain about a given probabilistic content. For example, it starts with a flat distribution over six options, assigning equal probability to each but with high uncertainty. After many samples it may move towards being reasonably certain that each option has probability 1/6th. I have no idea whether there are empirical examples of non-conceptual systems with this kind of sophistication. If there were, then we might treat the risk as a probabilistic content and the uncertainty as a probabilistic manner of representing that content. On the other hand, there is often good reason to treat uncertainty as a second order probabilistic content about one's own psychological states (Shea & Frith 2019). In this case the uncertainty could be a second order probabilistic content about the first order probabilistic content (or risk). So even with these two moving components the distinction between probabilistic contents and probabilistic manner is far from clear cut.

In sum, I think that Gross has posed an interesting and productive question. At this stage, however, we need to take seriously the possibility that no such distinction can be applied to non-conceptual representations, or that applying any tenable distinction to non-conceptual representations will produce an indeterminate verdict.

## Conclusion

It is a privilege to receive penetrating critiques from such respected reviewers. While I may not have been able to answer them all, I hope the debates show, at least, that subpersonal systems in cognitive science raise issues about the nature of content which are both tractable and philosophically interesting.

---

<sup>8</sup> We can still define the distance between a represented probability distribution and a categorical current world state (see fn. 6).

## Acknowledgements

I am very grateful to all three commentators, and to Ned Block who, together with Frances Egan, offered constructive criticism at an author-meets-critics session organised by Steven Gross at the 2019 APA Eastern meeting. Thanks also to John Morrison, Michael Rescorla and colleagues at the Institute of Philosophy for help with formulating my reply. This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No. 681422 (MetCogCon).

## References

- Block, N. (2018). If perception is probabilistic, why does it not seem probabilistic? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1755), 20170341.
- Egan, F. (2010). Computational models: A modest role for content. *Studies in History and Philosophy of Science Part A*, 41(3), 253-259.
- Egan, F. (2014). How to think about mental content. *Philosophical Studies*, 170(1), 115-135.
- Gallistel, C. R. (1990). *The Organization of Learning*. London / Cambridge MA: MIT Press.
- Gallistel, C. R., Craig, A. R., & Shahan, T. A. (2014). Temporal contingency. *Behav Processes*, 101, 89-96.
- Gallistel, C. R., & King, A. P. (2010). *Memory and the computational brain*: John Wiley & Sons.
- Gibbon, J., Berryman, R., & Thompson, R. L. (1974). Contingency Spaces and Measures in Classical and Instrumental Conditioning. *Journal of the Experimental Analysis of Behavior*, 21(3), 585-605.
- Godfrey-Smith, P. (2012). Signals: Evolution, Learning, and Information, by Brian Skyrms. *Mind*, 120(480), 1288-1297.
- Godfrey-Smith, P. (2014). Sender-receiver systems within and between organisms. *Philosophy of Science*, 81(5), 866-878.
- Gross, S., & Flombaum, J. (2017). Does perceptual consciousness overflow cognitive access? The challenge from probabilistic, hierarchical processes. *Mind & Language*, 32(3), 358-391.
- Martínez, M. (2019). Representations are Rate-Distortion Sweet Spots. *Philosophy of Science*, 86(5), 1214–1226.
- Moss, S. (2018). *Probabilistic knowledge*. Oxford: Oxford University Press.
- Shea, N. (2007). Content and its vehicles in connectionist systems. *Mind & Language*, 22(3), 246–269.
- Shea, N. (2015). Distinguishing Top-Down From Bottom-Up Effects. In S. Biggs, M. Matthen, & D. Stokes (Eds.), *Perception and Its Modalities* (pp. 73-91). Oxford: OUP.
- Shea, N., & Frith, C. D. (2019). The Global Workspace Needs Metacognition. *Trends in Cognitive Sciences*.

- Shea, N., Godfrey-Smith, P., & Cao, R. (2018). Content in Simple Signalling Systems. *British Journal for the Philosophy of Science*, 69, pp. 1009–1035.
- Skyrms, B. (2010). *Signals: Evolution, Learning, & Information*. Oxford / New York: OUP.
- Stocker, A., & Simoncelli, E. P. (2008). A Bayesian model of conditioned perception. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems*, vol. 20 (pp. 1409-1416). Cambridge, MA: MIT Press.