Perceptual Consciousness, Short-Term Memory, and Overflow: Replies to Beck, Orlandi and Franklin, and Phillips

Steven Gross and Jonathan Flombaum

Many thanks to our commentators—Jake Beck, Nico Orlandi and Aaron Franklin, and Ian Phillips—for their stimulating and challenging questions. Our paper was cast as a challenge to Ned Block's empirical arguments for the claim that perceptual consciousness "overflows" cognitive access. As a result, in places it slights the development of a positive alternative. Moreover, it sometimes clouds matters by offering multiple possibilities for development. The commentators press us for more detail. This is of course appropriate even if our central point was negative, since a challenge's force in part depends on there being a plausible alternative. They raise other worries as well. We won't do justice to them all, but will try at least to say something useful. There are some overlapping themes across their remarks; but rather than weave our replies together, which might hinder comparison with the commentaries, we organize them by author, taking up questions in the order raised.

It might be helpful, however, if we first say something about the paper's general strategy. Our aim was to show that current work in perception science provides ways to account for participants' performance in post-cueing tasks without hypothesizing a series of stores of declining capacity. These alternatives highlight instead the complexities involved in transitions from signal to representations to report—in particular, the various ways both encoding and retrieval provide opportunities for omission and error. We discussed, for example, how the path dependence of greedy algorithms can provide accurate encodings concerning one region at the cost of inaccurate or absent encodings elsewhere; and how retrieval in the form of sampling from a probabilistic representation can generate inaccurate reports. These explanations look rather different from models that posit an information bottleneck owing to stores of declining capacity. Whether, as recounted by us, they add up to a fully developed positive alternative depends in part on how detailed and complete one requires an alternative to be. But they are intended to indicate directions available in current work, the particulars of which will no doubt shift as research proceeds.

Suppose we have at least lessened the strength of Block's arguments for perceptual consciousness overflowing cognitive access. What *should* one say about perceptual consciousness? Though we entered some speculative remarks in our paper, here we will only underscore the interest of the particular form the question takes in light of the work we discuss: if the posited perceptual representations are probabilistic but the content of perceptual consciousness is not (as many, but not all, maintain), how are they related?

Beck

Beck raises three worries. First, he notes that the debate among competing conceptions of working memory is more open than our paper might suggest. Second, and third, he presents two ways we might be viewed as retaining overflow: owing to

1

information lost in the transition from a probabilistic representation to a discrete sampling therefrom, or owing to the number of items represented in one set of probabilistic representations in comparison to another. Before addressing Beck's worries, we provide one clarification that will also prove relevant in addressing the third worry.

In outlining our challenge to overflow hypotheses, Beck asks: "how is the visual system supposed to construct a representation of the cued row [in Sperling's post-cue paradigm] unless it has stored a representation of the full display," since "[o]nce the cue has arrived, the display is gone"? He notes one possible answer found in our discussion—that representations of lower-level features are stored, from which (in a manner affected by the cue) representations of higher-level features are generated—but notes as well our remark that we don't assume generally this sort of hierarchical representational dependence. Beck thus suggests that our more general argument turns on evidence for probabilistic representations in perception.

But this isn't quite right. Our argument against Block's use of Lamme and colleagues' results do turn on evidence for probabilistic representations in perception (as does one part of our response to Bronfman et al.). But, regarding Sperling, they play a role in just one of the various ways we develop our claim that the cue might post-dictively affect which letters (actually, alpha-numeric characters) get represented. The point of our remark concerning hierarchical representational dependencies is that our reply to Sperling goes through even for ground-level representations: representations that are not themselves dependent on representations of lower-level features. To see this, consider any case where the visual system generates ground-level representations from transduced signals—on the assumption that transduced signals, though they may in one of various senses carry information, are not themselves representational (Block 2014b, following Burge 2010, endorses such a position). Still, there might be post-dictive effects on the generation of those ground-level representations. The effect of a cue is just more signal. If it occurs within the post-dictive window, it can play a role in the generation of representations. It may be that this process involves probabilistic representations; but the general point doesn't require that.

Compare standard examples of post-diction, such as sound-induced visual bounce. Yes, the sound occurs *after* the disks have crossed. But that doesn't entail that *first* we represent the disks as not bouncing and *then* we represent the sound, and *then* we generate a conscious representation of the disks bouncing at the time of the sound. That *could* be. But it also could be that the signal produced by the disks and by the sound are processed together (even though the former is transduced before the latter) to yield the conscious representation, without a prior representation of non-bouncing disks. The point to be drawn from post-diction is that perception integrates information over a brief temporal window. That point can be naturally developed in probabilistic settings (perceptual systems in effect ask what out there *most likely* led to the signal accrued over this temporal window—and their answer may or may not preserve the temporal order of transduction), but the post-dictive point doesn't require it.

1. slot models vs. continuous resource models of visual working memory

Beck is certainly right that there is as of yet no consensus concerning the correct conception of visual working memory. The recent paper he highlights by Pratte et al. (2017) is an excellent example of the continuing debate. They advert to the fact that there

is less variability among perceptual representations of cardinal orientations (0°, 90°, 180°) than of oblique orientations, as well as a response bias towards the former. (One of us has done related work on color perception—see Bae et al., 2014; Bae et al. 2015.) The better performance of continuous resource models in previous studies, they argue, reflects an unwarranted advantage they have when such stimulus-specific variability is ignored; when the models incorporate it, slot models outperform continuous resource models, and hybrid models outperform them both.

Exploring various models is currently a central research focus, and no doubt will yield many intriguing results. It's worth noting, however, that, although "mixture models" like those found in Pratte et al. (2017) continue to be explored, other recent work points in a different direction. Whereas mixture models involve fitting parameters to a function and then making inferences about memory based on the parameter values obtained, the newer models involve simulating the activity of neural populations and then checking whether this activity could produce something like the experimentally-observed results. The newer models fit the data extremely well, and they are all (so far) either more obviously continuous-resource-like, or they have no resource limitations at all, with performance limits instead arising from the computations involved in encoding and retrieving data from the population code. Our paper mentioned Orhan and Ma's (2015) 'unmixing' model as one example of a model that does not involve any explicit resource limitation whatsoever. In their model, the more individual object signals are comingled in a population code, the less likely an accurate representation is retrieved—unmixed from the pool—for any of the objects. A more recent model without a resource constraint is found in Oberauer and Lin (2017), where memory limits arise entirely from interference among representations during retrieval. Schneegans and Bays (2017) explore a more continuous-resource-like model, wherein encoding and retrieval imprecision arise because of a continuous neural architecture that realizes feature-binding.

However the various models fare, the lack of current consensus applies in all directions. Thus, to the conditionalized thesis Beck offers us—"*if* continuous resource models prove correct, *then* we have reason to doubt the existence of overflow" (perhaps better: if a model that does not assume a capacity limit proves correct, then some of the main reasons offered for overflow can be questioned)—we would add the thesis that the slot model behind some of the arguments for overflow at least cannot be taken for granted. (We don't claim that Block simply takes it for granted—see Block 2008, pp. 300-1, though he doesn't happen to mention the way it is directly assumed in Lamme's groups' calculations.) Moreover, slot models currently in favor, because they incorporate probabilities, require capacity calculations distinct from those relied upon in the studies we discuss.

Finally, we note some wording in this part of Beck's discussion that we'd want to tweak—not to be picky, but to highlight a vehicle/content issue that's worth further consideration. Beck writes of a 'continuous resource of variable precision.' But we'd prefer to speak of a continuous resource doled out in a way that yields representational content of variable precision. The question of overflow concerns the number of items represented—a matter of content. Slot models assume a capacity limit regarding the number of items represented; continuous resource models do not. (This is not their only difference: crucially, continuous resource models assume probabilistic representations; at least simple slot models do not, though hybrid slot models do.) Talk of slots, however,

3

also suggests a limit on the number of *vehicles* in the store. As is often the case, there is thus the risk of confusing vehicle and content; and in the literature one does find talk of both types. But movement from one to the other need not simply involve a mistaken slide. Rather, it often seems there is a substantive claim or assumption being made about the relation of vehicle and content—viz., that, at least in this case, a memory store is content-limited in part because of its being vehicle-limited. Indeed, claims or assumptions relating the vehicles and contents are common in discussions of continuous resource models too. Thus, Bays (2015, p. 431), for example, characterizes slot models as "posit[ing] three or four independent memory slots, each holding a detailed representation of one visual object" and continuous resource models as "propos[ing] that a limited supply of a representational medium [e.g., amount of spiking activity] is continuously distributed between visual objects [presumably, *represented* visual objects]."

2. discretization and the loss of information

Beck suggests that there is a sense in which our view allows overflow, albeit one sufficiently different from Block's that it might be considered a reconception of it. The idea is that a probability distribution (or density—henceforth we leave this understood) contains more information than a sampling therefrom. We're happy to allow that there is a sense in which this is so—for instance, the sampling by itself contains no information concerning variance in the distribution. But it must be emphasized how distant this is from the claim Block defends.

First, whether this would support Block's claim that perceptual consciousness overflows cognitive access depends on how perceptual consciousness relates to the probability distributions. We discussed the possibility that perceptual consciousness is associated with a sampling. If so, the overflow Beck suggests occurs *prior* to perceptual consciousness. (We also mentioned the possibility of associating perceptual consciousness with represented probability distributions, citing Morrison 2016. See now also Denison 2017 and Morrison 2017.)

Second, it would depend as well on what counts as cognitive access. Sometimes entry into working memory is identified with, or considered sufficient for, cognitive access (Block 2007, p. 489). But the studies we cite concern probabilistic representations in visual working memory. So, one might see the representation of the probability distributions and the samplings as both occurring in cognition. (One might not, though, if one does not see VWM as part of the "global workspace" or otherwise a store entry into which suffices for cognitive access. For example, one might hypothesize perceptual working memories distinct from a central working memory. Incidentally, in our paper's original draft we invoked the idea that presence in working memory sufficed for cognitive access, but changed that, at a referee's suggestion, to the claim that it's necessary (Gross and Flombaum 2017, p. 359). Note that it's important that the claim concern working memory unrestricted, not more specifically visual working memory. Of course auditory representations needn't pass through visual working memory (perhaps multi-modal representations aside). But even restricting ourselves to visual representations, Block himself, in his interpretation of Bronfman et al.'s results, seems committed to the cognitive access of gist representations that bypass visual working memory.)

Finally, we should note that even if a probability distribution contains more information than a sampling therefrom, it doesn't follow that the information the sample loses is not accessed in some other way. Subjects' metacognitive judgments concerning their perceptions are taken by some to reflect such information (Kepecs and Mainen 2012).

3. overflow of the number of items represented

Beck also suggests that our view is compatible with there being more items represented in one set of probabilistic representations than another, so that the former can overflow the latter. The idea seems to be this: Assume that a sufficiently non-flat distribution represents things in accordance with the highest value it assigns (we ignore ties), and that a sufficiently flat distribution fails to represent anything. (Presumably there are also other senses in which they represent items or properties—for example, a distribution that assigns .5 to light from above and .5 to light from below presumably contains representations pertaining to light and location.) It is then possible that a transition from one set of probability distributions to another can amount to an overflow in this sense. For example, prior to the post-cue the visual system's probabilistic representations might represent 10-11 items, while after the cue they might represent 3-4 items, assuming the cue increases confidence in representations associated with the cued row while flattening the others.

We grant that merely hypothesizing probabilistic representations does not preclude this. But we have several reservations. First, our paper challenged Block's claim of evidence for overflow, not the very possibility of overflow. Second, and relatedly, what evidence is there that the visual system first represents in Beck's sense 10-11 letters and then, post-cue, only represents 3-4? Above, we clarified that, given post-diction, it's possible that no representations at all concerning specific letters are formed prior to the cue. (It could be, given the extensive training they receive, that subjects have a prior expectation of letters. And it may even be, as we mention, that the prior distribution may not be flat: perhaps not all letters are equi-probable. It's a further matter whether it would be sufficiently non-flat by Beck's lights.)

Further, it is possible that, by Beck's lights, the post-cue distribution *does* represent 10-11 letters. Sperling's capacity estimates are based on the number of *correct* reports, where subjects were required to place some letter in every location. If the reports were based on sampling from distributions, then it's possible that more than just 3-4 distributions were sufficiently non-flat to represent letters in Beck's sense, though only some resulted in veridical samplings. Presumably the probabilistic representations in the cued row are *more* non-flat than others and thus less open to sampling error, but there may be less non-flat representations (more subject to sampling error) in the other rows.

Moreover, one might question the assumptions behind Beck's suggestion. In the case of belief, how to relate credences and "full" (non-graded) beliefs is a disputed matter. Similarly, we may ask why probabilistic representations in perception should be assumed to represent in Beck's sense—i.e., be assumed to represent things as being in accordance with the highest probability assigned in the distribution. Indeed, our sampling hypothesis can be understood as a proposal for how to understand the relation between at least these sorts of probabilistic representations and discrete ones: the latter result from sampling the former. One might adopt this perspective as an alternative to Beck's

suggestion concerning what probabilistic representations represent in his special sense. Notice that they can diverge: one samples with a probability correlated with the probabilities assigned in the distribution, which allows for "error" (sometimes sampling a hypothesis not assigned the highest probability).

Finally, we underscore again that Block's overflow hypothesis is specifically about whether perceptual consciousness overflows cognitive access, not just whether there is one store that overflows another. As Beck notes, further argumentation would be required to relate any overflow claim here to Block's. In particular, if perceptual consciousness itself results from sampling, then the overflow from one probability distribution to another would not bear on Block's claim. (But, again, see Morrison's alternative.)

Orlandi and Franklin

Our paper suggests alternative interpretations for three experimental results that Block marshals in favor of his overflow claim—those due to Sperling, Lamme and colleagues, and Bronfman and colleagues. Orlandi and Franklin argue that, in each case, our alternatives face problems as well, sometimes problems parallel to those we raise for interpretations favoring overflow. We respond to each set of worries in turn.

1. letters

Orlandi and Franklin distinguish two ways we explain the Sperling results: one according to which the visual system fails to generate representations as of specific letters for some locations at all (henceforth, 'no representations'), and another according to which the representations it generates have too low a probability to make it to report (henceforth, 'low probability representations'). Orlandi and Franklin suggest it is unclear which explanation is ours. Our intention was to mark both as possibilities. In any event, Orlandi and Franklin argue that both have problems.

About the first ('no representations'), they present two worries: that the inferential aspect emphasized by our account becomes superfluous, and that it has difficulty accommodating subjects' reported phenomenology. The first worry, they hold, arises because, really, our argument just turns on the fact that processing takes time and attention can enhance processing, which doesn't require an inferential conception of visual processing. We wouldn't put it quite that way. Perceptual processing does take time, but our emphasis—following Phillips (2011a) and others—was more specifically on the fact that perceptual processes operate over a temporally extended signal, so that there is a post-dictive window within which a later stimulus (such as a cue) can affect the representation of an earlier stimulus. Still, it is true that one may accept this without endorsing an inferential account of perception. What such an endorsement might amount to depends on what one means by 'inferential.' We distanced ourselves from the term (Gross and Flombaum 2017, p. 366, fn. 9), since some require that representations be conceptual for transitions among them to be candidates for inference. Be that as it may, our fundamental disagreement with both Block and his previous critics is that Sperling's post-cues may affect, not just what is cognitively accessed, but what is represented in the first place—a point which might indeed be developed in various ways (as our paper's four examples underscore). That said, it is important to us that the point can be developed

in ways that are plausible, for example because they are in line with currently fruitful empirical inquiry. We are interested in more than mere logical possibility. Orlandi and Franklin suggest that our point could be developed in a non-inferential, purely associative manner. The plausibility of this suggestion depends in part on what makes a process purely associative. We note that the "dependencies" or "contingencies" our paper emphasizes could perhaps be described as associations, and that often those who do cast perceptual processes in inferential terms—for example, as a kind of Bayesian inference— allow that the algorithmic implementation might be associationist broadly construed (e.g., Griffiths et al. 2010).

Orlandi and Franklin's second worry for the 'no representations' view is that it has trouble explaining why subjects report seeing more than what they can remember. But we don't think this is so. First, it should be noted that the reports typically cited are anecdotal; and often it is not clear how to interpret them. For example, a person may say she saw twelve letters when she saw twelve items generically as letters, but not necessarily all twelve specifically as this or that letter (Gross and Flombaum 2017, p. 362 cites several replies to Block along these lines). In fact, in as yet unpublished work, Cova and colleagues found that fewer than half of their subjects presented with non-cued Sperling displays agreed that they saw all the letters; and, when probed further, most commonly selected descriptions like 'I saw some of the letters, but not all' or 'I saw all the letters but only some in detail.' Finally, Orlandi and Franklin suggest that the 'no representations' view predicts that subjects should report making a determination about they have seen based on remembered fragments. It's not clear to us that 'no representations' predicts that: why wouldn't subjects just say they aren't sure beyond the 4-5 they report, or that the other letters are 'vague' and 'blurry' (as Cova and colleagues report many of their subjects *do* say spontaneously in qualitative interviews)?

Orlandi and Franklin raise two worries as well for the 'low probability representations' view. The first is similar to Beck's suggestion that, really, we have reconceived overflow, instead of denying it. But perhaps a further remark here would be useful. Orlandi and Franklin say that, on this view, "the visual system represents a whole lot of letters, many more than the letters that are reported." But Sperling required subjects to report a letter for each location. So, perhaps what Orlandi and Franklin should say is that there would be more letters visually represented than *correctly* reported. It's in principle possible that probability distributions are generated for each letter, and that each is sampled, generating 12 discrete letter representations—12 cases of access. The flatter curves, however, would more likely lead to incorrect reports. (Cf. Gross and Flombaum 2017, pp. 384-5.)

The second worry for the 'low probability representations' view asks: "if it is troublesome to form discrete representations of letters, why is it any less troublesome to form probabilistic representations of letters?" The answer lies in the signal the visual system is dealt. Because the signal is noisy and ambiguous, the task of generating accurate representations as of specific letters may be too difficult, but it may well enable visual perception to generate a distribution of probabilities over the letters. Recall that the subjects are *expecting* letters, having received a high number of training trials (Gross and Flombaum 2017, p. 364). Arguably, the hypothesis space is already in place. It's figuring out which is where that's difficult. Compare: Someone is waving through the fog. It's not hard to know it's a human being. But it may be hard to know whether it's a man or

woman. Yet some retrievable evidence might give you more reason to think it's one rather than the other. It can be troublesome to form a (confident) "discrete" belief about gender, but less troublesome to form a probabilistic belief about it.

2. orientation

We point out that Lamme and colleagues' interpretation of their results assumes a slot conception of memory that is challenged by more recent continuous resource models; and we offer several, not necessarily exclusive explanations of subjects' change/no change reports in the various cueing conditions. Orlandi and Franklin raise a question about one of our suggestions: that "by redeploying attentional resources, the retro-cue can selectively protect corresponding representations of items in the first display from degradation and interference … [while] the post-cue [that comes with the second display] comes too late to have this effect" (Gross and Flombaum 2017, p. 377). They ask "why the post-cue would be too late, while the retro cue is not," suggesting the cues are "roughly equivalent when we consider [visual working] memory as a single store." But there are several reasons why they may not be equivalent. First, let us put to one side a reason ruled out by the experimental design. One might reason as follows: a memory representation, as our remark indicates, can deteriorate over time; also, the longer the interval, the greater the chance of it facing interference. The later the cue, therefore, the less likely it will succeed in successfully protecting the representation. This reasoning is right so far as it goes, and shows that adverting to the hypothesis of one memory store does not suffice to make Orlandi and Franklin's point (so, if they are right, it is not for the reason they give). But the comparative timing in Lamme and colleagues' retro-cue and post-cue conditions eliminates this explanation; for both cues occur 1000ms after the first display's offset. For that reason, we should not have spoken simply in terms of the post-cue coming too late. Nonetheless, a difference remains. The retro-cue occurs alone, while the post-cue occurs along with the second array. The presence of the second array itself may be a source of interference and cause deterioration of the memory representation. Indeed, Lamme and colleagues call their hypothesized intermediate memory store '*fragile* visual short term memory' precisely because it is said to be easily over-written. Further, and perhaps relatedly, the post-cue naturally directs attention to the relevant item in the second display. It may not be easy for one cue to direct attention *as well* to an item in memory.

3. color

Bronfman et al. (2014) – and Block (2014a) following them – argue that subjects' ability to judge color diversity, while also performing at capacity on a Sperling-like letter-identification task, supports "overflow." That color-diversity judgments come "for free"—that is, without causing letter identification to drop below capacity—is supposed to suggest, along with other results, that perceptual consciousness represents color-related features beyond visual working memory capacity—representations whose content presumably becomes available for report through some other pathway that by-passes visual working memory and its limited capacity.

Orlandi and Franklin suggest that these results pose a problem for the alternative conception of perceptual processing and visual working memory that we present, for it is unclear why, on our view, subjects should get color representations "for free." Of course,

we question whether subjects have perceptual experience as of specific colors for all or most of the letters. But we allow that they may have perceptual experience as of *some* sort of color-related feature—perhaps experience as of *color* (the genus) and experience as of high/low color diversity. Is it a problem for us that subjects seemingly get this "for free"? The problem can't be that subjects have a limited visual working memory capacity and yet are able to report color-diversity above chance, for our alternative does not posit a capacity limit on visual working memory. It seems the problem is rather supposed to be that the processing requires time or attentional resources in a way that our alternative can't readily handle. But it's not clear to us why. Gist perceptions, for example, are notoriously fast (Oliva 2005). And while it's been argued that they require attentional resources (Mack and Clarke 2012), it's not clear that the attentional resources available in this task are not sufficient. Though subjects are told that the letter-identification task is primary, they are also instructed that they will be asked about color-diversity and receive training to this end. Moreover, the attention required for a gist-based judgment of color-diversity is for features at a different scale than that required by the letter-identification task. It would need to be shown that the attentional resources compete, or do so sufficiently, to predict a difference in performance on color-diversity judgments in cued versus uncued rows (as per Orlandi and Franklin's final point)—especially as subjects were trained to associate those cues with the letter-identification task.

Phillips

Phillips asks four main questions: first, how do we regard our talk of representations and its relation to informational persistence; second, do we see recent work as superseding non-unitary models of iconic memory; third, using a point made by Matsukura and Hollingsworth (2011), might one might challenge Lamme and colleagues proposed intermediate store (fragile visual short term memory) while remaining neutral between slot and continuous resource models; and, fourth, to what bridging principles are we committed regarding information processing and phenomenal consciousness?

1. representations and informational persistence

Block holds that perceptual consciousness "overflows" cognitive access. This goes beyond the claim that subjects have perceptual experience that is not accessed. For perceptual consciousness and cognitive access could diverge without the former "overflowing" the latter (Gross and Flombaum 2017, p. 385). We followed Block in understanding the "overflow" claim in terms of comparative capacity. We followed him as well in understanding comparative capacity in terms of the number of items represented in distinct—presumably functionally individuated—representational stores (though Block doesn't happen to use 'store' as a nominal). We did so both because we are sympathetic to at least some of the assumptions behind these construals—for example, that perceptual experience is representational—and because it enabled us to frame our disagreement in terms that even Block would accept. (One wrinkle we should have added: One might distinguish the following two notions of overflow. First, the *capacity* of one store can be greater than that of another. Second, the *number of items in fact represented* in a store on some occasion can be greater than that of another. That a store has some capacity does not entail that it must always be used up. So, in the second

sense a store can overflow another of the same capacity, albeit for some reason other than that emphasized by Block.)

Phillips suggests that one might instead state the core disagreement in terms of information, which would add precision and not require waiting upon a worked-out theory of representation. But, he argues, if the debate is so formulated, the question arises whether we can avoid commitment to a rich store after all, in addition to a later visual memory store.

The term 'information' is itself used in many ways, so more detail would be needed to assess whether it provides a gain in precision. (Of course, it's not expected that Phillips develop the idea in a brief commentary.) Phillips refers to "some trace of the stimulus which persists after offset for the cue to allow its selective encoding." This is indeed a less demanding notion than we had in mind, encompassing, as it does, even the transduced signal. (It is perhaps sufficiently weak as to follow constitutively from what perception is, at least in the veridical case, that such information is contained even in the transduced signal.) We can allow that there are senses, including this, in which there may be more information at an earlier stage of processing than a later stage. Of course, as Phillips no doubt would agree, this by itself won't help Block, since he would not associate perceptual consciousness with the signal or any other non-representational state. Does recasting the debate in informational terms sharpen it or just reveal further fault lines? The debate concerns whether one perceptually experiences more than one cognitively accesses. If there is good reason to maintain that perceptual experience is representational, then the debate ought not to be recast in other terms. Our apparent disagreement with some other challenges to the overflow claim—such as Phillips'—might then turn on substantive disagreements about perceptual experience rather than "a certain looseness in talk of representations."

Two further comments. First, contrary to Phillips' suggestion, we did not commit ourselves to there being only one visual store. We discussed both that possibility and the possibility that there are successive visual stores of the same capacity. That said, his argument that we might be committed to there being at least two visual stores, if correct, would of course preclude the possibility of there being only one. Second, we explored both the possibility that vision might fail to form representations of some letters and the possibility that it may form probabilistic representations of all of them. It is not clear (absent details concerning the relevant notion of information) why the second possibility would not to be consistent with the later visual store containing in some sense the same amount of information concerning specific letters as the earlier store.

2. supersession of non-unitary models of sensory memory?

Phillips also asks if the single-store view we consider supersedes earlier models that challenge the idea of a unitary sensory memory. The short "answer" is that we need to think more about how to relate these results and debates to our suggestion. Part of the complexity of this question stems from Phillips' point that it's not obvious what constitutes a store—in particular, what functional properties might individuate them and what sort of nesting and dynamic changes they might allow. In places, we invoked hierarchical perceptual processing, allowing that this may involve not just hierarchical dependencies among representations, but also a temporal hierarchy in their generation. One could argue that this temporal progression itself suggests a series of stores. At

another point (Gross and Flombaum 2017, p. 383), we allow time-sequenced representations within a single store, albeit there we had in mind that this was in virtue of represented temporal content. (The distinction between temporal properties of representations and temporal representational content receives some discussion in Gross (forthcoming).) The question of what constitutes a store and how they are individuated—and whether this is a useful construct in the first place—bears also on Phillips' last question concerning how information-processing models are related to phenomenal consciousness. As Phillips' notes, Irvine (2011) argues that a multiplicity of sensory stores—as opposed to a unified iconic memory—significantly complicates and perhaps undermines such questions. Unclarity as to what a store even is complicates them further. But Phillips presents another question for us to deal with first.

3. alternative challenge to Lamme and colleagues
We argued that Lamme and colleague's capacity calculation assumed a slot model of visual working memory that can be challenged. Phillips asks what we think of another challenge, one that applies even if one upholds a slot model. Actually, he asks what "precisely [we] think one needs to commit to in order to resist the postulation of fragile VSTM." Rather than address that formulation head on, perhaps we may say that there may be various interesting challenges to fragile VSTM or to arguments for it, and it's worthwhile to identify them all and their varying commitments. Because they may be committed in different ways, they may be open to different counter-replies. ("Your honor, my client doesn't even own a dog. And, anyway, his dog doesn't have any teeth.")

The challenge comes from Matsukura and Hollingworth (2011). They suggest that calculating Cowan's K on a post-cue change-detection task may over-estimate capacity, because subjects will be incented to forget uncued items, whereas Cowan's capacity formula assumes that subjects attempt to encode and retain all items. Now, Lamme and colleagues' subjects may well have attempted to encode all items into memory, since they didn't know where a post-cue might point. (It's possible that some subjects, finding the task difficult, adopted a strategy of just picking a region and hoping for the best, like a goalie on a penalty kick who just chooses a side in advance. That would be a different challenge, which we'll bracket.) Does the post-cue's incentive to forget non-cued items then cause problems for Cowan's calculation and its retention assumption, assuming a slot model? It's unclear it does—at least with simple slot models. Presumably, by the time of the cue, items have either made it into memory or not. The cue cannot cause anything to be added to memory. Does forgetting non-cued items affect performance, on a simple slot model? Let's assume that the subject makes no correspondence errors, of the sort we suggested could alternatively account for at least some of Lamme and colleagues' results. Let's likewise assume that the memory representations are all-or-nothing and don't interfere with one another. How does dropping *other* items from memory improve performance in reporting whether the cued item has changed? The model provides no reason it should. (Orlandi and Franklin's question would also have to be addressed, though perhaps we have already suggested how: Why wouldn't an even later cue, one that comes along with the second array, likewise incent one to forget non-cued items, with similar results?)

There may be good answers to these questions. Certainly there are more sophisticated capacity-limited models that allow for variable precision. But they would

need to be considered in detail, while assessing whether the mechanisms hypothesized fit the data as well as the continuous resource alternative. Our paper's second proposed way of accounting for Lamme and colleagues' results bears a family resemblance to Matsukura and Hollingworth's suggestion, but is developed in a probabilistic manner (Gross and Flombaum 2017, pp. 376-7).

4. bridging information processing and phenomenal consciousness

Our paper engaged in some speculative rumination concerning how perceptual consciousness might relate to representational stores hypothesized by a probabilistic account of visual working memory. Phillips asks us to consider the possibility there is no mapping—well, no *simple* mapping—between stores and phenomenology of any kind and that it's a "fundamental mistake" to think computational models could overturn phenomenologically-based claims concerning perceptual consciousness. He asks in particular what "bridge commitments [we] think constrain the relationship between a story told at the information processing level without mentioning phenomenal consciousness and claims about phenomenal consciousness."

We are certainly open to the possibility that the relations are complex, perhaps ultimately even inscrutable to beings like us; Phillips is right to broaden the space of options. But, for that reason, we do not think in terms of bridge commitments. We think that it is important to explore bridge *hypotheses* and that this is the most likely way we will make whatever progress we are capable of making on these difficult topics, even if in the end the progress (as it were) is negative. As Phillips points out, from the perspective of the alternative viewpoint he floats, it is "obscure how to connect disputes regarding the modeling of short-term visual memory and disputes concerning the contents of visual consciousness." (We are put in mind of Lee's (2014, p. 5) reaction to what could seem an attempt by Phillips (2011b, 2014) to detach temporal experience from its apparent neurophysiological and computational underpinnings—or at least to complexify their relations: "I find this view hard to understand.") In inquiry, we try to find a foothold where we can, and sometimes (shifting metaphors) light is shed, sometimes not. While we don't rule out Phillips' possibility, we also don't rule out the possibility of phenomenological claims on these matters being overturned by plausible empirical evidence that meshes with plausible larger theories.

Our commentators have given us much to think about, and, as promised, some of it we will need to think about much more. Thanks again!

References

Bae, G., Olkkonen, M., Allred, S., Wilson, C., and Flombaum, J. 2014. Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision* 14, p. 7.

Bae, G., Olkkonen, M., Allred, S., and Flombaum, J. 2015. Why some colors appear more memorable than others: a model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General* 144, pp. 744-63.

Bays, P. 2015. Spikes not slots: noise in neural populations limits working memory. *Trends in Cognitive Sciences* 19, pp. 431-8.

Beck, J. 2017. Some worries about the "no-overflow" interpretation of post-stimulus cueing effects. *The Brains Blog*, this symposium.

Block, N. 2007a. Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavorial and Brain Sciences* 30, pp.481-99.

Block, N. 2008. Consciousness and cognitive access. *Proceedings of the Aristotelian Society* 108, pp. 289-317.

Block, N. 2014a. Rich conscious attention outside focal attention. *Trends in Cognitive Sciences* 18, pp. 445-7.

Block, N. 2014b. Seeing-as in the light of vision science. *Philosophy and Phenomenological Research* 89, pp. 560-72.

Bronfman, Z., Brezis, N., Jacobson, H., and Usher, M. 2014. We see more than we can report: 'cost-free' color phenomenality outside focal attention. *Psychological Science* 25, pp. 1394-403.

Burge, T. 2010. *Origins of Objectivity*. Oxford: Oxford University Press.

Cova, F., Gaillard, M., and Kammerer, F. (unpublished manuscript). Subjective data and the overflow argument. University of Geneva.

Denison, R. 2017. Precision, not confidence, describes the uncertainty of perceptual experience. *Analytic Philosophy* 58, pp. 58-70.

Griffiths, T., Chater, N., Kemp, C., Perfors, A., and Tenenbaum, J. 2010. Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences* 14, pp. 357-64.

Gross, S. forthcoming. Perception and temporal representation. *Pacific Philosophical Quarterly*.

Gross, S., and Flombaum, J. 2017. Does perceptual consciousness overflow cognitive access? The challenge from probabilistic, hierarchical processes. *Mind & Language* 32, pp. 358-91.

Irvine, E. 2011. Rich experience and sensory memory. *Philosophical Psychology*. 24, pp. 159-76.

Kepecs, A., and Mainen, Z. 2012. A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367, pp. 1322-37.

Lee, G. 2014. Temporal experience and the temporal structure of experience. *Philosophers' Imprint* 14, pp. 1-21.

Mack, A., and Clarke, J. 2012. Gist perception requires attention. *Visual Cognition* 20, pp. 300-27.

Matsukura, M., and Hollingworth, A. 2011. Does visual short-term memory have a high-capacity stage? *Psychonomic Bulletin & Review* 18, pp. 1098-104.

Morrison, J. 2016. Perceptual confidence. *Analytic Philosophy*, 57, pp. 15-48.

Morrison, J. 2017. Perceptual confidence and categorization. *Analytic Philosophy* 58, pp. 71-85.

Oberauer, K., and Lin, H. 2017. An interference model of visual working memory. *Psychological Review* 124, 21-59.

Oliva, A. 2005. Gist of the scene. In L. Itti, G. Rees, and J.K. Tsotsos (eds.), *Encyclopedia of Neurobiology of Attention*. San Diego: Elsevier, pp. 251-6.

Orhan, A., and Ma, W. 2015. Neural population coding of multiple stimuli. *The Journal of Neuroscience* 35, pp. 3825-41.

Orlandi, N., and Franklin, A. 2017. Getting clear on the challenge. *The Brains Blog*, this symposium.

Phillips, I. 2011a. Perception and iconic memory: what Sperling doesn't show. *Mind & Language* 26, pp. 281-311.

Phillips, I. 2011b. Indiscriminability and experience of change. *The Philosophical Quarterly* 61, pp. 808-27.

Phillips, I. 2014. The temporal structure of experience. In D. Lloyd and V. Arstila (eds.), *Subjective Time: The Philosophy, Psychology, and Neuroscience of Temporality*. Cambridge: MIT Press, pp. 139-58.

Phillips, I. 2017. Commentary on Steven Gross and Jonathan Flombaum's 'Does perceptual consciousness overflow cognitive access? The challenge from probabilistic, hierarchical processes.' *The Brains Blog*, this symposium.

Pratte, M., Park, Y., Rademaker, R., and Tong, F. 2017. Accounting for stimulus Specific variation in precision reveals a discrete capacity limit in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance* 43, pp. 6-17.

Schneegans, S., and Bays, P. 2017. Neural architecture for feature binding in visual working memory *Journal of Neuroscience* 37, pp. 3913-25.

Sperling, G. 1960. The information available in brief visual presentations. *Psychological Monographs*, 74.

Vandenbroucke, A., Sligte, I., and Lamme, V. 2011. Manipulations of attention dissociate fragile Visual Short-Term Memory from Visual Working Memory. *Neuropsychologia* 49, pp. 1559-68.