

18

Introspection and Unrevisability: Replies to Commentaries

Murat Aydede & Donald D. Price

We are grateful to our commentators for their very useful and insightful comments. We are delighted to see that there is much agreement on key issues between us and the commentators, which might be an indication for a sea change in the scientific study of pain and other types of conscious experiences. This makes the issues addressed all the more important and urgent.

Gallagher and Overgaard

Gallagher and Overgaard's commentary has two parts. The first and shorter part is on introspection in general. The second, longer part contains a very useful survey of positions similar to ours along with a critical internal comparison among them. We have almost nothing to say about the second part except to express our gratitude for such an enormously useful comparison of our proposal to the converging work of others — enormously useful to us, at least, since we were not familiar with many of them. The present form of the view proposed here has its origins in the work of Price and Barrell (1980, 1984, 1985) as well as Barrell and Barrell (1975). We are of course delighted to learn that there has been such a convergence and looking forward to learning more about it while savoring the realization that such a growing convergence might be an exciting sign of what is to come in the study of human consciousness.

In the first part of their commentary, however, Gallagher and Overgaard argue, contrary to our view, that not every form of first-person access to one's conscious experience and its features is introspective. Introspection, in other words, is not the only first-person way of knowing our experiences. They criticize us for not making the distinction, even though, as they themselves remark, our proposal demands only the introspection as they characterize it. Namely: "...introspection is understood as a second-order cognitive act that thematizes first-order phenomenal experience, and makes that experience the object of reflection." So, in a sense, their criticism is clearly incidental and not directed against our main proposal. Nevertheless, we want to make two points on this.

First, in the target article we recommended introspection as a passive noticing of what is or has just taken place in experience and reporting the content of experience in a manner that faithfully reflects the experience (first person, present tense). We certainly proposed that the participants reflect on the common factors or elements that were present in multiple experiences of a given phenomenon, such as anger or performance anxiety, but only after the experiences have taken place and have been reported. This listing of common elements was then to be followed by an analysis of the necessary and sufficient elements of the type of experience in question. In principle, this is similar to the phenomenologists search for structural invariances. The quoted definitions listed by Gallagher and Overgaard appear not to clearly reflect the distinction between the act of observing the contents of experience and the act of analyzing their nature. We think this is an important distinction regardless of how introspection is defined.

Secondly, we are very skeptical as to whether there is any first-person direct access to one's own experiences that doesn't involve a second order mental state whose content is about the experience (or an aspect of it) thus accessed — as indeed the very notion of access suggests.

Gallagher and Overgaard talk about our having "... an implicit, non-introspective, pre-reflective self-awareness of our own experience." They continue: "At the same time that I see the light, I know that I see the light. This knowledge of seeing the light is not based on reflectively or introspectively turning our attention to our own experience. It is rather built into our experience as an essential part of it, and it is precisely that which defines our experience as conscious experience." It is not clear to us how we should understand this talk of knowledge about our experience being built into the experience itself. Experience is one thing, the knowledge of it is another, especially reportable knowledge. Our awareness (knowledge) of experiences is not mediated by the very same experiences. To be sure, experiences are typically representational mediators through which or with which we become aware of what those experiences sensorially represent. But it is totally mysterious to us how an experience can mediate our awareness of it through the very same experience itself — unless, by this, we just mean to make the rather obvious remark that they are conscious states (but this is a different matter). If we look at the history of philosophy, we see a conception of introspection (especially in Kant, but also in British empiricists like Lock) as internal perception. Just as ordinary external perception has its object and is distinct from it, introspection was meant to have other (immediately available) mental states as its objects while being distinct from them. This much, we assume, is essential for introspection and not controversial. But we also think that it is almost analytic that any first-person (direct, non-inferential) knowledge that we can report about our experiences is introspective in just this minimal and rather traditional sense (possibly including the historical phenomenological tradition). Note that we didn't claim that for an experience to be conscious it has to be introspected (or, even introspectable). But these are deep waters that need extended discussion. Given that their criticism is not directed at how we propose to use 'introspection' where this is understood in the sense they and we both agree upon, we may safely put this issue aside.

Gallagher and Overgaard complain that we seem to trivialize the use of introspection when we say that introspection is and has been indispensable for psychophysics (and other similar experimental paradigms). They complain that in the sense in which we use 'introspective report', any sincere utterance counts as introspective reporting (among other things) since they indicate that the speaker had an intention, say, to communicate. We think that this is not the right way to look at the issue. Although some psychophysical experiments are designed in terms of detecting the presence or absence of stimuli, many are designed in terms of estimating magnitudes and qualities of sensory experiences. In the case of magnitude judgments, this requires that the subject notice and represent sensory magnitude using a rating method or representing the perceived magnitude of one type of sensation by using another sensory modality (e.g., adjusting a sound to represent the intensity of a pain). This task requires the subject to notice what is present in their experience and represent what is present in the form of words or non-verbal responses. The main bulk of psychophysics is not concerned with the relation of stimulus parameters and any response on the part of the subjects. Most often and crucially, it is concerned with the relation between the stimuli and sensations — conscious sensations. Thus, magnitude judgments inevitably reflect a simple form of introspection.

Importantly, in the case of pain (and other similar bodily sensations), reports of pain are introspective reports partly because pain is less easily linked to stimulus properties than are other sensory modalities (cf. IASP definition of 'pain'). Notwithstanding the confused folk practice of locating pain in body parts, pains are experiences in the first instance. Reporting pain is therefore reporting experiences (among other things). If this is done in a first-person way,

genuine pain reports are introspective in our sense. For instance, subjects are often not asked to just detect the noxious stimuli in classical psychophysical sense but rather are asked to determine when feeling the stimulus becomes painful, i.e., pain. Thus, pain thresholds, for example, could not have been established by simply asking the subjects when they detect certain kinds of stimuli.

Finally, at some point in their commentary, Gallagher and Overgaard question our suggestion that scientists should use themselves as subjects, similar to Nahmias' statement that we are not explicit enough about why we think experimenters should first test their ideas on themselves. Clearly, our rationale for this position needs further clarification and justification. Gallagher and Overgaard remind us that investigators have certain hypotheses and results that they hope to find and thus are more likely to be biased as subjects. We fail to see why this should be a concern particularly for our proposal, especially since evaluations of biases of either investigators or subjects are by no means routinely made. Investigators become susceptible to bias the moment they form their hypotheses or design their experiments and there are multiple ways they can transfer their biases to their subjects. Biases have numerous origins, including the past or existing literature, imagining how something works, and vested interest in supporting ones past results. Far from being a problem for our position, involving the investigators themselves as subjects has the potential of eliminating or minimizing the impact of such biases within the formal framework of our proposal. Indeed in the experiential method we propose, the participation of investigators as subjects offers a means whereby bias can be recognized and dealt with more effectively. If the subject matter of a study is that of a particular type of experience, such as a type of pain or an emotion, then wouldn't it be better for the investigators to encounter these phenomena in their own direct experience as the basis for formulating hypotheses and designing experiments as opposed to only relying on published accounts of others or their imagination? The former approach reflects being objective about the subjective, whereas the latter approach reflects being subjective about the subjective. The experiential approach may be especially effective when there are multiple investigators/subjects. The exploration and results can be compared across participants and checked for possible biases. It is important to note in this regard that judging and bias can be part of what is passively noticed by the participant, as pointed out in our article.

Robert D'Amico

D'Amico thinks that there is a radical epistemic asymmetry between first-person introspective data and third-person empirical data: while the latter are revisable and refineable, the former are not — not even in principle. He thinks that the asymmetry is not marked merely by degree but rather it's the manifestation of a radical difference between two conceptual schemes. Because of this, they cannot be put into service in the way our proposed methodology requires; in other words our methodology is deeply confused by treating these two types of data as if they were on a par — conceptually. He writes: "We can not imagine, I suggest, that the experimenter could be led by some persistent or even robust brain imaging results to the conclusion that such and such an experience that he or she is now having is not in fact a pain experience after all; or that the diffuse pain felt in that testing was in fact a sharp pain after all, given the experiment's third-person results." This would be imaginable if our methodological proposal were sound. Evidently, D'Amico thinks that it is logically inconceivable that someone competent with the relevant mental vocabulary and conceptual repertoire and familiar with the normal/ordinary range of sensory experiences, taking his time and upon careful reflection, can be mistaken about

whether or not he is having a certain sensation of a familiar kind, or whether he is having this sensation rather than some other ordinarily easily distinguished sensation. Although D'Amico doesn't put the point in modal terms, we believe he would not object to our putting his unrevisability claim as a metaphysical/logical impossibility claim: under optimal conditions, competent subjects cannot be mistaken in their core judgments about their own current sensations above threshold levels. By "core judgment" we mean identifying their own sensations in terms of their general kind and above-threshold spatio-temporal properties. The cases suggested by D'Amico's own examples serve well in this regard. Here is another rather extreme example — to pump the relevant modal intuition. Suppose under optimal psychophysical conditions, I take a good bite of a ripe cantaloupe melon, chew it for 10 seconds, and sincerely misidentify the taste as a good quality dark chocolate taste. The intuition we are invited to share here is that necessarily if I indeed sincerely (to my own great surprise) identify the taste as a dark chocolate taste, then it must have tasted to me so. It is unimaginable (thus, metaphysically impossible) that I have the melon taste but sincerely identify it as the taste of dark chocolate. In short, one can certainly make subtle mistakes about one's sensations at threshold levels, but no gross mistakes under optimal conditions are possible — there is simply no logical room here. Call this unrevisability claim the Cartesian Transparency Thesis regarding one's Current Sensations (CTT_{cs}). We think CTT_{cs} is false. Thus we reject the radical unrevisability claim that D'Amico presents as constituting an essential barrier to the kind of methodology we propose.

The issues surrounding this traditional and venerable thesis are deep and many. We cannot hope to address all the questions raised by our rejection. However, we do want to provide some support for why such a rejection is not crazy. First, we want to point out that there is a distinction between having a sensation or experience and making a judgment about an experience one is currently having. However tight the connection may be, it is clear that they are distinct events that do not metaphysically/logically necessitate each other. Many animals or even young children may plausibly be said to have experiences without being capable of making judgments about their experiences. It is plausible to attribute to some animals sensations without attributing capability of making judgment at all (let alone judgments about their own sensations). Furthermore, the psychoneural mechanisms subserving these capacities are probably functionally distinct. The capacity to have sensations may not necessitate the capacity to have concepts and use them in making judgments — even though having a sensory system may be nomologically necessary for having concepts at all. Thus, insofar as CTT_{cs} is a thesis about the revisability of one's judgments about one's current sensations and insofar as making a judgment is a distinct activity from having a sensation, it should in principle be metaphysically/logically possible that these judgments are revisable in a way that makes CTT_{cs} false. This is a sort of transcendental argument against CTT_{cs}, but one might legitimately ask whether we have any positive conception of how revisability of sensation judgments is possible. We think we can come close enough to make positive sense of this possibility, but explaining how would go beyond the scope of this reply. However, we don't think that our rejection of CTT_{cs} depends on our offering a positive conception. Moreover, we think that there are very good (principled) reasons why we are typically having difficulty in forming a first-person positive conception of such cases. Nevertheless, we would like to present some empirical cases which it is plausible to describe as cases where a mismatch occurs between one's conscious experience and one's judgments about these in a way that puts the burden on the defenders of CTT_{cs} to show that they don't falsify CTT_{cs}.

For example, repeated encounters with the same stimulus object may lead to different sensations depending on several psychological factors. Coghill's example of the thermal grill illusion (see commentary by Coghill) offers an example of how a stimulus that evokes a complex sensation produces very different types of subjective reports depending on how the experimenter instructs the subjects to attend and the methods used to report sensations. In the case of his example, if the same subjects had been exposed to the same thermal grill in the two experiments described by Coghill, it is very likely they would have revised or refined their own reports of how they experience the thermal grill in a context that is not influenced by biases fostered by the experimenter.

More radical cases may be certain forms of anosognosia — in particular visual anosognosia (Anton-Babinski syndrome) where people suffering from this disorder typically claim, indeed insist, that they can see, offer visual descriptions of what they claim to see, or attempt to move around, all the while being technically blind, i.e., not capable of receiving and processing visual information around their immediate surrounding in a way that matches their verbal behavior and their behavior. When asked about why they bump into objects or when the accuracy of their reports is challenged, they often confabulate, sometimes wildly, but rarely accept their blindness. There are competing theories that could potentially explain what is happening in visual and other anosognosias. While some postulate as part of their explanation that these patients are in fact having hallucinatory sensations (unrelated to their actual surrounding reality), others attempt to explain them as involving complex disconnection problems among various subsystems of which the disconnection of the conceptual/verbal system from others including various sensory visual subsystems plays an important role. Our point here is not that the latter explanation should be preferred, but rather the fact that an explanatory hypothesis of this sort could be offered shows that the rejection of CTT_{cs} where radical error of this sort is logically ruled out is not so crazy after all.

Other cases may involve some forms of agnosias, especially forms of associative visual agnosia, where low level vision of details is often preserved but the patient cannot recognize the visual stimulus. One way to describe such a syndrome is by saying that although the patient can see every low level visual detail (say, lines, edges, colors, light and color gradations, etc.) more or less just like a normal subject can see, the patient cannot see the object presented in this detail as what it is. What would the visual phenomenology of such patients be like? Despite good visual acuity, the visual world of ordinary objects seems to disappear from the visual phenomenology of such patients. These patients are not blind, their entire visual field preserves some qualitative phenomenology, in fact in some sense exactly like the phenomenology of the normal subjects, but they just can't see the apple, the lighter, match box, the pencil, the cup, the plate, etc., presented in this phenomenology. These patients often are unable to visually imagine such objects or remember what they looked like — despite the fact that they have near perfect non-visual concepts of these objects: when they touch, taste, hear, or smell them, often times they can recognize them easily. It is as if almost all the visual information has been erased from the mental files (concepts) they keep for object kinds. Sometimes some patients can figure out what the visually presented objects must be by painstakingly reasoning out from the low level visual cues they can see.

This case doesn't directly bear on the possible falsity of a radical unrevisability thesis about one's current sensations. But it shows how recognitional judgments in a given sensory modality can be impaired while the phenomenology of (early) sensory processing in that modality is preserved. (We are assuming that, for instance, visual object recognition is partly a

modality-specific conceptual capacity involving classifying different visual stimuli under a visual concept.) Anosognosia is a deficit that attaches to various neuropsychological deficits, such as blindness, hemianopsia, hemiplegia, amnesia and others. We are not aware of an actual case of associative visual agnosia with anosognosia. Such a case may well be empirically impossible, but then again, maybe not. If such a case is empirically possible, then we would have a person who cannot visually identify objects but insists that she can — despite having complete (low-level) visual acuity and phenomenology. A person in this condition would judge, for example, “it appears to be a cup” when in fact nothing visually appears to be a cup to her. Again, our point here is not that there may be actual cases of this sort (for all we know, there may be). But rather, certain empirical hypotheses (e.g., disassociation between sensory systems and conceptual information pick up or categorization or verbalization mechanisms) naturally suggest themselves that cannot be put aside a priori. If these hypotheses turn out to be better than others in explaining such deficits, CTT_{cs} would be in serious trouble even as a nomological necessity claim — we would have actual counterexamples. But we believe that we have already shown that we have possible counterexamples. Hence CTT_{cs} is false as ordinarily defended — as a metaphysical/logical necessity claim. And this is enough to rebut D’Amico’s main criticism that there is a radical asymmetry in kind between first-person and third-person modes of epistemic access to data, which shows according to D’Amico that we have two radically different (logically incongruent) conceptual schemes involved in these two kinds of access that cannot be experimentally juxtaposed in the way our methodological proposal requires.

Finally we cannot help but point to the actual practice of scores of experimental neuropsychologists who have been using first-person data fruitfully for decades and now increasingly in brain imaging studies with great success and amazing results that have sometimes immediate practical and clinical consequences.

Robert Coghill and Eddy Nahmias

We don’t really have much to say about Coghill’s and Nahmias’ supportive commentaries. They expand on some of the things we touch upon in our paper briefly, and so in this regard they are very helpful and complementary — for which we are grateful. We have only a few minor points to make.

In the beginning of his paper, Coghill writes: “In other words, while being burnt with a 51°C stimulus, one is not aware of the thermal qualities of the stimulator, but is simply experiencing pain. In contrast, while one is being stimulated with a 40°C warm stimulus, one attributes the sensation of warmth to the external object (i.e., the stimulator).” In our paper, we emphasize the differences between pain and other more obviously representational perceptual experiences. But we don’t mean to deny that pain experiences are representational, they represent actual or potential tissue damage and its various features. Of course, pain experiences do more than just represent. Also, we don’t think it is plausible to say that in having a thermal pain caused by, say, 51°C stimulus, one is not aware of thermal qualities. It may be that in having the pain experience one is aware of these thermal qualities while registering them as (potentially) damaging the skin area the stimulus is applied to.

We concede Nahmias’ point that not all dualist positions require lawlike psychophysical correlations. However, we are puzzled about why he thinks that Cartesian interactionist substance dualism suggests that no such correlations will not be found. If psychophysical causation (causal interaction in both directions) requires psychophysical covering laws, then Cartesian dualism entails that there are lawlike correlations between some properties instantiated

in a physical body and some properties instantiated in a non-physical substance (soul) “attached” to that body. But perhaps Nahmias has in mind the high variability of pain experiences under the influence of other mental states. Such a variability would certainly make things difficult, but probably not impossible. Nevertheless, we agree with Nahmias’ insightful observation “that [our] methodology offers a way to put pressure on other metaphysical theories by discovering the correlations that those theories predict should not be found.”

We also agree with Nahmias’ claim that our “methodology may work differently depending on what introspection is — and if there are different types of introspection, it may depend on what type the experimenters and subjects use.” In the target article, we didn’t have space to discuss the difficult issue of what introspection consists in, and so we left its conception at an intuitive level. But this issue needs addressing especially in the light of our positive methodological proposal. One of us (MA) has a detailed theory about the basic mechanisms of experiential introspection and its ontogenesis, according to which just having an experience of pain is a different activity/event than introspecting it. Introspection requires cognitive (conceptual) capacities — see our reply to D’Amico above, and Aydede and Güzeldere, 2005, forthcoming, for more details.)

References

- Aydede, Murat, and G. Güzeldere (2005). “Cognitive Architecture, Concepts, and Introspection: An Information-theoretic Solution to the Problem of Phenomenal Consciousness,” to appear in *Noûs*.
- Aydede, Murat, and G. Güzeldere (forthcoming). *Information, Experience, and Concepts: A Naturalistic Theory of Phenomenal Consciousness and its Introspection* (Oxford, UK: Oxford University Press).
- Barrell, J.J. and Barrell, J.E. (1975). “A self-directed approach for a science of human experience,” *Journal of Phenomenological Psychology*, Fall, pp. 63–73.
- Price, D.D. and Barrell, J.J. (1980). “An experiential approach with quantitative methods: A research paradigm,” *Journal of Humanistic Psychology*, 20(3): 75–95.
- Price, D.D., and Barrell, J.J. (1984). “Some general laws of human emotion: Interrelationships between intensities of desire, expectations, and emotional feeling,” *J. Personality* 52: 389–409.
- Price, D.D., Barrell, J. E., and Barrell, J.J. (1985). “A quantitative-experiential analysis of human emotions,” *Motivation and Emotion*, 9: 19–38.