BEHAVIORAL AND BRAIN SCIENCES (2001) **24,** 187–193
Printed in the United States of America

Continuing Commentary

Commentary on **Jeffrey A. Gray (1995). The contents of consciousness: A neuropsychological conjecture. BBS 18:659–722.**

Abstract of the original article: Drawing on previous models of anxiety, intermediate memory, the positive symptoms of schizophrenia, and goal-directed behaviour, a neuropsychological hypothesis is proposed for the generation of the contents of consciousness. It is suggested that these correspond to the outputs of a comparator that, on a moment-by-moment basis, compares the current state of the organism's perceptual world with a predicted state. An outline is given of the information-processing functions of the comparator system and of the neural systems which mediate them. The hypothesis appears to be able to account for a number of key features of the contents of consciousness. However, it is argued that neither this nor any existing comparable hypothesis is yet able to explain why the brain should generate conscious experience of any kind at all.

Facing the hard question

Włodzisław Duch
Department of Computer Methods, UMK, Torun, Poland.
duch@phys.uni.torun.pl

Abstract: The following questions are considered: Why is it difficult to create a theory of consciousness? What are the contents of consciousness? What kind of theory is acceptable as transparent? and, What is the value of conscious experience?

Gray (1995) claims that we are still a long way from creating a transparent theory of consciousness and a completely new kind of theory is needed. Neither in his target article nor in the commentaries is much progress made on the Hard Question: How to create a scientific, causal theory of the links between consciousness and brain-and behavior? Gray's article leaves a rather pessimistic impression. Is the Hard Question really so hopelessly difficult? and What are the reasons for this lack of progress?

Lesion studies have taught us a lot about the localization of various mental functions but consciousness is much more robust than such cognitive abilities as recognition of words or faces. As Newman (1995) pointed out in his commentary, destruction of several regions of the brain, notably the RAS (reticular formation of the brain stem) and ILC (intralaminar complex of the thalamus), induces coma. Extensive damage to subcortical structures, as described by Gray (1995) in his response to the commentaries, may lead to alterations of conscious experience. Many drugs also produce various changes in conscious experience, from enhancing certain qualia to producing a zombie-like state. Since conscious experience does not seem to depend on localizable neural tissue, we must assume that a number of structures are necessary to generate it, and some of the relevant circuits have been presented by Gray (1995) in Figure R1.

Consciousness is a particularly difficult subject to study because experiments on animals are of limited usefulness and there is little data relating human conscious experience to brain damage. We cannot cut off all memories of things red and then see how this will influence the qualia of looking at red. In addition, as Freeman (1995) has shown, the same stimulus and same behavior do not imply similar neural activity (see also Skarda & Freeman 1995). The approximately invariant entrainment of smaller groups of neurons may be embedded in chaotic activity of larger neuronal groups and could therefore be difficult to find. It is not just the activity or lack thereof, but also the proper synchronization of activities (entrainment) of several brain structures that is important for conscious experience.

Although the empirical difficulties are serious, suppose that some day we are able to determine how the proper entrainment of thalamo-cortical reverberation correlates with subjective, conscious experiences. It may even be possible to selectively "switch off" some neural nuclei for a limited time and observe the effect of such changes of brain processes on perception and qualia. Will the knowledge obtained in this way constitute a brute correlation of brain and mind events or will it lead to a transparent theory that Gray is hoping for? This is the central question considered here.

Is consciousness dependent on information processing or on brain states? Information processing in the brain is ultimately done by molecules (Black 1994); it is therefore based on the real physical states of very complex matter. No amount of information processing will change a simulated vibration into a real vibration. I do not see any reason to believe that qualia and consciousness may arise out of pure information processing. Conscious experience depends on activation of real biological matter. Experiences recalled from the memory are similar, but not quite identical to the brain activations of the original experiences and thus the corresponding qualia are somewhat different. The subjective, first person perspective is about my states of the brain and body while the objective, third person perspective is about the description of these states. As Rachlin (1995) wrote in his commentary on Gray, sensation is located in the functional interaction of the whole body and the environment. Digital processing of information misses not just the causal properties of neurons, as Searle (1980) points out, but it fails to reproduce the physical states of the gray matter. The evolution of these physical states may be described as information processing, but description is not reality. The states of the reticulo-thalamo-cortical (RTC) feedback loop give rise to a particular experience called consciousness. Can we understand how this happens?

The content of consciousness may result from the output of a number of comparators, of which the subicular comparator may be the most important. However, when I reflect on my own experience – despite problems of introspective psychology, I believe that in this case we really need a better phenomenology and books like Varela et al.'s (1993) are a good start in this direction – my subjective feeling of being conscious does not depend on novelty of stimuli but rather on arousal, that is, processes of attention mediated by the RAS system. Zen monks practicing concentration for many hours a day report a strong feeling of being conscious although they have no novel stimuli that could mismatch their ex-

© 2001 Cambridge University Press 0140-525X/00 $12.50
https://doi.org/10.1017/S0140525X01213934 Published online by Cambridge University Press

pectations. In this case it is the RAS system itself which seems to maintain the high degree of vigilance and conscious feeling. It is possible that such states of concentration are just well synchronized (focused) neural RTC states leading to strong qualia. The essence of conscious experiences does not seem to lie in the evolution from one mind object to the other. It lies rather in the exploration of a single mutimodal object: a thought, a sound, a visual scene, each having many features inducing complex brain/body reactions, leading to the specific physical states of the whole organism, states dependent on individual history, hence subjective states. Bodily reactions in anxiety are not just symptoms but essential parts of the experience: symptoms and causes are not separable, hence somatic therapy may have strong psychological consequences.

The theory of consciousness that I find satisfactory is based on (1) physical states of the brain, evolving according to internal dynamics created by genetic as well as environmental factors, and on (2) a correlation between these physical states and subjectively reported qualia. The question "Why should the brain create conscious experience?" does not seem to be more reasonable to me than the question "Why should two gases, such as hydrogen and oxygen, create water?" Indeed, in the early days of chemistry, this was something very difficult to accept. We can now predict some properties of water starting from quantum mechanics, and we should be able to predict (from the third person perspective) the existence of qualia expressed as subtle behavior arising from the comparison of the stimuli with memorized experiences. We learn at school that water is a mixture of two gases and accept this as a fact. Why should learning that our mental experiences are an emergent property of the brain be harder to accept? Theory will never reduce "being it," or the first person perspective, to "describing it," or the third person perspective (in this sense, the mystery of conscious will never go away, as Dennett [1995] wrote in his commentary).

Such a theory should also account for the survival valve of consciousness. The ability to empathize with others (contrary to what Gray (1995) claims in his reply, sect. R2) does not require other conscious minds and may be accumulated gradually. The internal dynamics of the physical states of the brain, from which conscious experiences emerge, allow escape from the animal's "here and now." The evolutionary advantage of consciousness lies in the ability to avoid inflexible behavior patterns (based mostly on genetic learning) that animals follow. Consciousness and intelligence (adaptability to a complex environment) are inseparable.

ACKNOWLEDGMENT

Support from the Polish Committee for Scientific Research, grant 8T11F 00308, is gratefully acknowledged.

The contents of consciousness: From C to shining C++

Michael H. Joseph[a] and Samuel R. H. Joseph[b]

[a]*Department of Psychology, Institute of Psychiatry, University of London, London SE5 8AF, United Kingdom;* [b]*Centre for Cognitive Science, University of Edinburgh, Edinburgh EH8 9LW, United Kingdom.*

spjtmhj@iop.kclf.ac.uk **srhj@cogsci.ed.ac.uk**

Abstract: We suggest that consciousness (C) should be addressed as a multilevel concept. We can provisionally identify at least three, rather than two, levels: Gray's system should relate at least to the lowest of these three levels. Although it is unlikely to be possible to develop a behavioural test for C, it is possible to speculate as to the evolutionary advantages offered by C and how C evolved through succeeding levels. Disturbances in the relationships between the levels of C could underlie mental illness, especially schizophrenia.

It appears to us that many of the problems in discussing the biology of consciousness (C) so clearly enunciated by Gray (1995), arise from the implicit assumption that C is a unitary phenomenon which is possessed by some nervous systems, and perhaps by some conceivable neural networks, but not others. The problem of how C arose (evolved) and what advantages it conferred is more easily addressed if we adopt a mutilevel concept of C. In fact Gray has already accepted two levels of C, in that he defines his conjecture on the neuropsychology of C as underlying the primary awareness of Jackendoff (1987). As quoted by Gray, primary awareness extends to "the perceived world, qualia, body sensations, proprioception, mental images, dreams, internal speech, hallucination, etc." This seems to encompass a wide range of rather different phenomena, some of which fall outside the scope of Gray's hypothesis. We would prefer to separate these into at least two levels: primary C (perceptual scene including the organism's position within it (egocentric space), environmental maps (allocentric space), limited knowledge of and use of past experiences), and secondary C (abstract knowledge, recordable and generalisable experience, thinking, rehearsal, recapitulation, mentally trying solutions, thinking about a subject when it is not present). Whereas primary C is present in many animal species, secondary C would be present in only a few animal species other than man, and to varying degrees. Tertiary C, corresponding to Jackendoff's reflective awareness (appreciation of other mental states, beliefs, value systems) would be found, as far as we are aware at present, only in humans.

Since many elements in Gray's conjecture are drawn from animal experiments, especially those in the rat, it would appear that animals should have, at least in some form, the type of C being described. Thus, Gray's system would represent the substrate of primary C in our scheme. This level of C could have evolved as a means of handling the vast influx of information flooding into the nervous system of higher animals, partitioning processing appropriately between conscious and unconscious levels, and accordingly improving their chances of learning from experience and surviving. (We may note in passing that if C is truly an emergent property [from increasing complexity], then there is actually no *requirement* that C itself confer a survival value.) This level of C does not require simultaneous emergence of C in many individuals of the same species to confer a survival advantage.

Secondary C could be seen as evolving from primary C through the ability of the animal to create and consider environmental maps from other spatio-temporal location (e.g., "what would I see if I were over there"). Secondary (and tertiary) C (e.g., "imagine what someone else would see if they were over there," etc.) could then be instantiated in successive elaboration of the same neural circuits.

Alternatively, they could use different neural circuits which are more elaborated in higher animals, for example, cortico-cortical circuits. If these higher levels did use the comparator system postulated by Gray for the analysis of match/mismatch by simulating the effects of perceptual input using other areas of the cortex (a plausible biological route in that it removes the necessity of producing a separate analytical structure), it would imply some form of "tagging" so that the animal would be able to distinguish between imagined situations and the current situation. Errors in such a tagging system could result in confusion between thought and reality.

The survival value of the ability to perform an operation simulating intended actions is apparent, in that it allows an individual to predict a negative or positive outcome in advance, without undergoing the potential hazard of performing the action. Tertiary C would evolve through the development of symbolic representation, which facilitates thinking, into language; one of the most immediate survival advantages would be to facilitate cooperation in hunting and foraging. One might even be so Machiavellian as to conjecture that the capacity to appreciate others' mental states, for which awareness of one's own is a pre-requisite, evolved because it conferred the advantage that we could more effectively manipulate the mental state of others, to control their behaviour. Again this ability would not have to emerge simultaneously in all indi-

https://doi.org/10.1017/S0140525X01213934 Published online by Cambridge University Press

viduals of a species. The development of language and tertiary C then permit cultural evolution (in practice, the "inheritance" of acquired characteristics, impossible in Darwinian evolution) to occur.

As Gray points out, our own C is a datum; that of others, even of our own species, is an inference based on their behaviour. This is really the force of the Turing test for intelligent machines; if a machine behaves, in every detail, in such a way as to make us unsure whether or not we are interacting with a conscious human, then we have no alternative test of C to apply. The same argument applies to animals: if they behave in such a way that C provides a convincing account, then we can treat them as being conscious, but this cannot be demonstrated unequivocally. In some sense, C does not alter behaviour; it will always be *possible* to elaborate other explanations for an animal's (or indeed another human's) behaviour based on reinforcement learning, or the copying of another individual. C provides us with an efficient description of their behaviour, which Occam's razor leads us to entertain, but does not lead to a behavioural test for the presence of C.

It has been convincingly demonstrated that autism is associated with a failure to develop an appreciation of the mental states of others (Frith 1989), which we have allocated here to tertiary C. While it may be tempting to think of schizophrenic symptoms as arising in a similar way, it would appear that for some symptoms, for example, paranoia, a theory of mind was, on the contrary, a prerequisite. How can you suspect others of having evil intentions towards you unless you know that they can have intentions? We would speculate rather that the successful control of successively higher levels of C, without spiraling off into the blue yonder of higher and higher levels of abstractions, depends upon a continual anchoring of higher levels of C in lower levels, a continual checking that meta-statements also fit with common sense and with "real" perceptions at a lower level. The implication is that the positive symptoms of schizophrenia might arise from failures to integrate and coordinate between the different levels of C. Conversely, the ability to loosen these links *in a controlled way* could underlie creativity, and perhaps explain the link, so often commented upon, between schizophrenia and artistic creativity.

At a very simple level, we can see the role of dopamine in latent inhibition, discussed by Gray, in mediating between a lower level of C (the current stimulus contingencies), and a higher level of C (something "known" about the CS on the basis of prior experience). Thus dopamine disruption of LI may indeed be a model for the effects of dopamine on the disturbances of integration between lower and higher levels of C which more plausibly underlie the strange beliefs of schizophrenia.

ACKNOWLEDGMENT

MHJ is a member of the MRC External Scientific staff; SRHJ is the holder of an MRC scholarship.

Must all action halt during sensorimotor mismatch?

Daniel M. Merfeld

Jenks Vestibular Physiology Laboratory, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA 02114.

dan_merfeld@meei.harvard.edu **www.jvl.meei.harvard.edu/jvpl**

Abstract: Gray's target article presents a model of consciousness that includes several ideas similar to those developed over the past century to explain how sensorimotor information is interpreted by the nervous system. This commentary discusses these ideas and introduces some additional hypotheses, also derived from sensorimotor investigations, that might help improve Gray's model.

Gray (1995) has derived an interesting, testable, and perhaps important model of consciousness from his earlier models of anxiety and schizophrenia. As he points out, "at least part of the neural activity that gives rise to conscious experience should remain closely tied to the different perceptual systems themselves." Despite recognizing this, Gray appears to neglect many relevant concepts developed by sensorimotor physiologists. Some of these concepts appear to support the conceptual framework of Gray's approach, while at the same time suggesting potential enhancements to his model. For example, Gray suggests that all motor programs must halt in the presence of certain mismatches. Other models (more congruent with some experimental evidence) take a different approach, using continuous (or nearly continuous) feedback to help minimize the mismatch.

Evidence suggests that at least three sources of information play crucial roles in sensorimotor processing (e.g., Bridgeman et al. 1994): (1) inflow from sensory receptors (*feedback*), (2) copies of the efferent signals (*feedforward*), and (3) experience interpreting structural sensory cues (e.g., rotations of the eyes lead to predictable changes in the retinal projections of the stationary external world). As pointed out by Grusser (1994), evidence of some of these concepts can be traced at least as far back as the early seventeenth century. More recently, Sperry (1950) and von Holst and Mittelstaedt (1950) helped formalize these concepts when they independently suggested that motor commands must leave an image of themselves (*efference copy*) somewhere in the central nervous system that is then compared to the afference elicited by the movement (*reafference*). It was soon recognized that the efference copy and reafference could not simply be compared, since one is a motor command and the other is a sensory cue (Hein & Held 1961; Held 1961). To solve this problem, Held developed two conceptual elements, a *Comparator* and *Correlation Storage.* The comparator in Held's schema appears indistinguishable from that presented by Gray. Held (1961) pointed out that under this scheme "the re-afferent signal is compared (in the Comparator) with a signal selected from the Correlation Storage by the monitored efferent signal. The Correlation Storage acts as a kind of

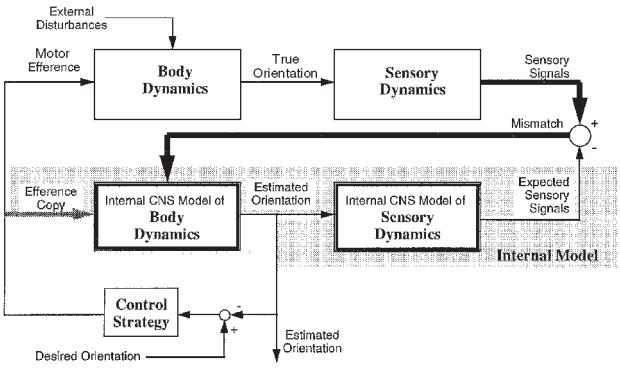


Figure 1 (Merfeld). Block diagram of the internal representation model. The primary input to this model is *desired orientation,* which when compared to the *estimated orientation* yields *motor efference* via a *control strategy.* These motor commands are filtered by the *body dynamics* (e.g., muscle dynamics, limb inertia, etc.) to yield the *true orientation,* which is measured by the sensory systems with their associated *sensory dynamics* to yield *sensory signals.* In parallel with the real-world body dynamics and sensory dynamics, a second neural pathway exists that includes an *internal representation of the body dynamics* and *an internal representation of the sensory dynamics.* Copies of the efferent commands (*efference copy*) are processed by these internal representations to yield the *expected sensory signals,* which when compared to the sensory signals yield an error (mismatch). This error is fed back to the internal representation of body dynamics to help minimize the difference between the *estimated orientation* and true orientation. (Modified from Merfeld 1995b.)

https://doi.org/10.1017/S0140525X01213934 Published online by Cambridge University Press

memory which retains traces of previous combinations of concurrent efferent and re-afferent signals."

Recently, a mathematical representation of these ideas has been developed to help explain the process by which the nervous system interprets sensorimotor information (Merfeld 1995b; Merfeld et al. 1993). The underlying concept is that the nervous system knows something about the dynamics of its sensory and motor systems and uses this knowledge to develop an *internal representation* of these dynamics. The development of this internal representation is guided by experience interpreting sensorimotor cues, including correlations between various cues. Figure 1 shows a representation of this model, which includes each of the three information sources discussed previously. In brief, (1) the thick black arrows (*sensory signals*) represent the feedback pathway by which the sensory systems influence the estimates of the current states, (2) the thick gray arrow (*efference copy*) represents the feedforward paths which help predict what the sensors will measure based on any planned action, and (3) the highlighted boxes (*internal CNS models*) represent neural processes that help the nervous system interpret structural sensorimotor cues based on previous experience. These internal representations are hypothesized to match as closely as possible the dynamics associated with the sensory and motor systems. This model also includes a comparator, similar to that discussed by Gray, by which the sensory signals are compared to expected sensory signals.

The similarities between this model with an internal representation and Gray's model of consciousness are somewhat surprising. Analogous to Gray's description of his model, this model (1) takes in sensory information; (2) interprets this information based on motor actions; (3) makes use of learned correlations between sensory stimuli; (4) makes use of learned correlations between motor actions and sensory stimuli; (5) from these sources predicts the expected state of the world; and (6) compares the predicted sensory signals with the actual sensory signals.

However, this model differs from Gray's somewhat in the way that it handles mismatches. For example, it does not (7) decide whether there is a mismatch between the expected and actual states of the world; and (8) bring the current programs to a halt in case of mismatch. Instead, it uses the difference (mismatch) between the expected and actual sensory signals as an error signal to guide the estimated state back toward the actual state even when the signals do not match (as opposed to Gray's step (8), which repeats steps (1) through (7) only when there is a match). Analogously, this model suggests that motor actions also include continuous (or nearly continuous) adjustments, with the difference between the estimated state and the desired state continuously affecting the current motor commands. This model works extremely well for modeling the processes by which the nervous system combines multi-sensory information during complex motion stimuli (Merfeld 1995a; 1995b; Merfeld et al. 1993) and for predicting eye movements (a motor response). Since consciousness must remain tethered to the sensory systems, this simple change might help improve Gray's model.

This adjustment appears necessary because normal human subjects can be aware of conflicting sensory information under a variety of conditions and can still generate appropriate motor responses. For example, when upright subjects observe full-field visual display that is rotating in roll (i.e., rotating about an axis aligned with the subject's line of sight), they perceive a continuous sense of angular velocity aligned with the roll motion of the visual field (*vection*). Simultaneously, they perceive a tilt relative to gravity that eventually reaches a steady-state value of approximately 10 degrees (Dichgans et al. 1972). When queried, subjects consciously recognize that it is not possible to rotate about an axis that is perpendicular to gravity and be tilted statically (i.e., the sensation of tilt should change depending upon the velocity and direction of the sensed rotation). Yet each of these conflicting sensations is so robust that subjects maintain these perceptual states even after recognizing the conflict. Furthermore, subjects are able to elicit appropriate motor responses during this disturbing mismatch. For example, voluntary and reflexive eye movements are observed even after the subjects recognize the conflicting sensations, and subjects can freely move their arms, hands, heads, and so on. (Many subjects can even lift a foot from the floor, but this is somewhat tricky because of the balance problem introduced by the sensation of tilt in the absence of actual tilt!) Hence continuous control of motion along with conscious awareness of conflicting perceptions of motion and orientation are present even during this type of sensorimotor mismatch. Continuous, or nearly continuous, feedback could help alleviate the apparent discrepancy between these findings and Gray's model. For example, might it be possible to use a mismatch to activate Gray's hypothesized Behavioral Inhibition System while also continuously adjusting behavior using feedback of the mismatch?

The approach suggested by this model might also help clear up another small difficulty with Gray's model. In section 3, Gray suggests that the predicted states of the world are compared to the actual states. Of course, as acknowledged by Gray, there is no way to know the actual state of the real world. We can only estimate the actual state based on sensorimotor cues. The present model gets around this problem by comparing the *actual sensory signals to the expected sensory signals* and using this difference in a feedback loop (under most conditions) to guide the estimated state toward the actual state.

While parsimony is not a principle of science, one strength of this model is its ability to continuously embody different sources of sensorimotor information, even when the information from these sources appears incongruent. Knowledge gained from this approach might help guide the development of Gray's model by eliminating, or at least reducing, the need to halt motor programs when sensorimotor mismatches occurs.

ACKNOWLEDGMENT

The work was supported by NASA grant 199-16-11-54 and NIH/NIDCD grant DC03066.

Motivating consciousness

Ian Vine

Department of Interdisciplinary Human Studies, University of Bradford, Bradford BD7 1DP, United Kingdom. **i.vine@bradford.ac.uk**

Abstract: Gray's account of a brain mechanism for generating the contents of consciousness is incomplete. Adaptive advantages of conscious functioning need to be sought within the first-person affective sensation motivating flexibly goal-directed actions, as in Humphrey's sensory feedback theory.

Resolving crucial explanatory difficulties with Gray's (1995) neuropsychological comparator theory of consciousness could hinge on explicitly integrating such models with sensation, motivation, and affective processes.

1. The "hard" problems. Functional or neural accounts of how complex brains generate conscious awareness have yet to show how phenomenological features supervening on neural activity are necessary for appropriate information-processing or behavioural outcomes (see symposium in *Journal of Consciousness Studies*, vols. 2(3) to 4(1), 1995–1997). Gray admits that his theory is no exception. The corollary is that we cannot grasp how awareness could have evolved at all, except improbably as a non-causal epiphenomenon.

2. Hints toward solution. Without stipulating the precise philosophical form of a solution, I take the "subjective character" (Nagel 1974) of conscious mental outcomes not to be replaceable by computational functions or neurophysiological structures and states. The privacy of qualia precludes capturing their agent-centred "feel" in objective, third-person descriptions of representational processes. As Velmans (1995, p. 703) argued, conscious

https://doi.org/10.1017/S0140525X01213934 Published online by Cambridge University Press

phenomena "do not seem reducible to either a physical or a functional state of the brain." We should "start to take consciousness *in the form that we normally experience it* seriously."

Informational events with a phenomenal aspect over and above their material one can be reported verbally and socially shared. Predicting or communicating about subjective experience may be amenable to an adaptive story, but it presupposes higher-order reflexive- and self-awareness (Humphrey 1984; 1993) rather than revealing why primary awareness first evolved.

If primitive phenomenology partly cognate with our own awareness has a phylogeny, it should inform attacks on the hard problems. Gray's comparator system may generate conscious thoughts, but it remains focused on "the primary role of consciousness in reflecting the external world" (1995, p. 708); Gray understands awareness "as a monitoring process" (1995, p. 672). Velman's arguments tell against the idea that Gray's representational mentality (causally) *requires* any subjectivity.

Bodily pains and pleasures may both predate awareness generated by complex matching processes and may hold the key to understanding the hard problems. Pain awareness seems to *amplify* the adaptive impact of relevant neural activation (Flanagan 1992). Evidence of neural pathologies is lacking for the rare cases of congenital insensitivity to pain that seriously jeopardizes survival (Horn & Munafò 1997). Moreover, pain is perhaps more exhaustively defined by its subjective properties than are other sensitivities. Gray acknowledges the oddity of saying "pain is a form of monitoring" (1995, p. 672) – although he agrees that inputs from pain receptors may interrupt ongoing comparator processes, intrude into awareness after initiating emergency reactions, then gain functional utility by signalling that a motor program was faulty (cf. Toates 1995). But if all monitoring is computational, why do pains need to be felt as qualia?

Humphrey (1993) noted how inter-penetration of sensation and perception is typically strong, especially for our distal senses. Yet he boldly conjectures that both kinds of processing arose in parallel rather than being serially connected. Direct perceptual processing uses sensory inputs but excludes conscious affect; it evolved to represent useful information about external regularities. Proximal senses primarily represent the state of one's body itself – sensations being inherently ego-centred and affectively valenced for the experiencer. Humphrey sees them as evolutionary residues of primitive aversive or appetitive "wriggles," whereby primitive organisms deal with stimuli at their bodily periphery. Once sensorimotor responsive control had migrated to an integrative brain system, sensory qualia eventually emerged as the activity of reverberating circuits, sustaining momentary stimulation. Primary consciousness *is* having sensations that constitute an extended "subjective present," while the enduring neural pulses represent "sentiments," or valenced response dispositions. These remain poised to restore a necessary conative push as soon as afferent information has been matched intelligently against repertoires of stored plans and values. Sensations could be said already to embody incipient *intentions* to implement voluntary action choices.

3. Affective sensations as paradigmatics? Amit's (1995) review of the legacy of Hebbian theories about cortical cell-assembly loops which sustain afferent excitations from brief stimuli in short-term memory has some resonance with Humphrey's theory. So does Edelman's (1994) identification of a perceptually oriented primary consciousness with activity in thalamo-cortical "re-entrant loops." Gray (1995) mentions Humphrey's theory only in passing. His observations on such ideas about specifically sensory representations of organismic rather than external-world states would be instructive.

Irrespective of the neural instantiations involved, Humphrey reinforces the case for starting analyses of consciousness with inherently motivated awareness, most evident with basic bodily sensation like pain. Authors like DeLancey (1996) share the conviction that affective awareness requires new research attention. Although cognitive science may believe that computational models do not leave the information-processor lost in thought, we know that without affect intelligent action can be substantially impaired. The structures on which Gray focussed – especially in the modified (1995) model which links in thalamo-cortical circuitry – surely provide a rich playground for seeking the motivating role of affect in turning perceptual comparisons into actions. Elaboration I have gestured toward might take us closer to the core of the hard problems if they could identify a role for sensation in the affective/conative economy of intelligently adaptable organisms.

Author's Response

No easy answers to hard or easy questions

Jeffrey Gray

Institute of Psychiatry, King's College London, London SE5 8AF, United Kingdom. **spjtjag@iop.kcl.ac.uk**

Abstract: What makes conscious experiences necessary for information processing or behaviour (no one knows)? Would it be easier first to divide consciousness into different levels (probably not)? Is consciousness tied to information processing or brain states (no one knows)? Would the target article's comparator be improved by adding a continuously adjusting feedback (probably not)?

Vine comments that neither the model of the contents of consciousness's advanced in the target article (in which this point is indeed admitted) nor any other existing theory has yet shown "how phenomenological features supervening on neural activity are necessary for appropriate information-processing or behavioural outcomes." This is an accurate statement of one aspect of the hard problem of consciousness. Vine suggests that a solution to the problem might be closer to hand if one started, not from the notion that conscious experience reflects, in perception, the external world (as proposed in the target article), but from Humphrey's (1983; 1993) notion of primitive sensations. These are said to be closely linked to motivationally valenced bodily states (involving pain and pleasure) and to their associated action tendencies. Like Humphrey, Vine proposes, furthermore, that the initial evolution of conscious experience occurred in the context of these bodily sensations and action tendencies. While I accept that this may be a plausible position from which to start on the search for aspects of behaviour and information-processing that are most intimately related to the evolution of conscious experience, I fail to see that the defect in my own model, identified by Vine and noted above, would be any less evident if one were to change the centre-piece from perceptual representations of the world to sensation-plus-action tendencies. It is just as true of the latter that there is no understanding of why these "phenomenological features supervening on neural activity are necessary for appropriate information-processing or behavioural outcomes." What could it be about motivationally valenced bodily sensations that requires the evolution of qualia, and how do such sensations differ from perceptual representations in requiring them?

To the extent that there are any data that bear upon this choice, they do not support the Vine-Humphrey position. There is now considerable evidence that action tendencies

https://doi.org/10.1017/S0140525X01213934 Published online by Cambridge University Press

in the evolved human case (e.g., reaching out for a glass of water to quench a motivationally valenced thirsty feeling in the throat) are mediated by a system (the so-called "dorsal stream") that operates without concomitant conscious awareness; and this system is dissociable (as shown in a wide spectrum of neuropsychological disorders) from the perceptual representations that themselves appear introspectively to be the sine qua non of conscious experience (Milner & Goodale 1995; Weiskrantz 1997).

Like Vine, **Duch** comments on an aspect of the hard problem: "Is consciousness dependent on information processing or on brain states?" Of these two possibilities, he opts for the latter, but based upon abstract arguments only. Such arguments can take one only so far; witness the fact that, also using arguments of this kind but at book length, Chalmers (1996) has come to exactly the opposite conclusion. Difficult though the task will be, the time has come to bring such issues into the laboratory. Until relevant experimental evidence can be brought to bear, both possibilities, that information processing or brain states – and indeed the further possibilities that both information processing and brain states, or even neither of them – are necessary and/or sufficient for consciousness are likely to remain open. I have discussed elsewhere (Gray et al. 1997b; Gray 1999) some experiments, currently under way, on "coloured hearing synaesthesia" that may perhaps throw some light upon this issue.

Duch sees the hard problem as being less hard than I do. He sees no essential difference between the question (the hard one): "Why should the brain create conscious experience?" and this other one: "Why should two gases, such as hydrogen and oxygen, create water?" The proposed parallel between these two questions is, however, misleading. I am no chemist. But I believe it to be the case that this science has gone well beyond the initial discovery that, as a matter of fact, water can be decomposed into hydrogen and oxygen. Chemists can now give a rather precise account of the properties of water in terms of the properties of the molecules of hydrogen and oxygen that combine to make it, together with an account of the atomic structure of hydrogen and oxygen themselves (plus some further even more microscopic levels of explanation). In the case of the brain and conscious experience, we do not at present have any conception of what such a mechanistic account of how the latter derives from the former would look like. When we have such a concept, at least in outline, a solution to the hard problem will finally be in sight. We are not there yet; but I see no reason to lower our scientific standards in advance of trying to apply them to the problem of consciousness.

Joseph & Joseph propose a hierarchy of consciousness divided into three levels. While this tripartite division may eventually be shown to be both useful and accurate, it does not seem likely to aid in the solution to the hard problem of consciousness addressed by Vine and Duch. If we could grasp what it is about brain and behaviour that required the evolution of the simplest level of conscious experience (represented perhaps by the sensation of pain, as stressed by Vine), I suspect that the rest would then fit rapidly into place. Such a fit might very well take the form, advocated by Joseph & Joseph, "that the successful control of successively higher levels of C [consciousness]. . . depends upon a continual anchoring of higher levels of C in lower levels." But, in the absence of any scientific understanding of what the lowest level of consciousness consists in, it is premature to speculate about higher levels. Indeed, despite the plausibility of Joseph & Joseph's suggested distinctions, it may yet turn out that the very metaphor of "lower" and "higher" levels of consciousness is misleading.

Merfeld addresses a much more specific feature of the model of the contents of consciousness proposed in the target article, namely, that these consist in the outputs of a comparator system charged with the function of determining which components of the current description of the external world (as computed in thalamo-cortical perceptual systems) are as predicted (on the basis of the previous state of the world, past regularities of experience under similar conditions, and the subject's current motor program) and which are not so predicted (or, better, under-predicted and by what degree). As Merfeld points out, this model is very similar to comparator models used in other branches of psychology and physiology, and in particular in the analysis of sensorimotor function. The similarity is no accident, and I should perhaps have acknowledged more explicitly this parentage to the model. Merfeld further comments that standard models used to account for sensorimotor function utilise continuous feedback so as to minimise discrepancies ("mismatch") between expected and actual input to the comparator, whereas the model I have proposed allows for motor programs to be brought to a halt in the event of significant mismatch. This is indeed a major difference between my comparator model and those used to analyse sensorimotor integration. It is not one, however, that is unmotivated.

I first proposed the comparator model set out in the target article in the context of a theory of the neuropsychology of anxiety (Gray 1982a; 1982b). The paradigmatic situation in which this emotion arises is that of conflict between an approach and a passive avoidance tendency, in which one tendency or the other must eventually dominate (for a detailed analysis of the concept of conflict in this and other, different, situations, see Gray & McNaughton 2000). Within this context, the function of the comparator is to scan the environment for stimuli that are either associated with negative outcomes (punishment or frustrative nonreward) that may arise from the current motor program or which represent a radical departure from expectation and may therefore constitute a source of danger. If the threat evaluated from either of these sources is sufficiently great, then it is imperative to interrupt the ongoing motor program, so as both to prevent further approach into danger and to permit the adoption of alternative, active avoidance strategies. Improvement in the precision with which the motor program attains the goal of the approach tendency (the normal use of mismatch in the type of feedback circuitry envisaged by **Merfeld**) would, by itself, still leave the animal in a situation that might be unacceptably dangerous or uncertain. It is for this reason that the comparator in this model needs to be given the capacity to operate an output of behavioural inhibition, interrupting ongoing motor programs.

The target article further exploited this comparator model in an effort to find a solution to one aspect of the hard problem of consciousness, namely, the fact that in many, perhaps most, instances conscious awareness of stimuli in the external world comes too late to affect behaviour directed toward or away from these stimuli (McCrone 1999; Velmans 1991). This lateness of concious experience is accounted for, within the model, by the hypothesis that the

https://doi.org/10.1017/S0140525X01213934 Published online by Cambridge University Press

comparator function that is relevant to conscious experience occurs after incoming stimuli have already been used to guide ongoing behaviour. Conscious perception then acts as a late error-detection device (this being the cognitive function corresponding to the emotion of anxiety). Thus, **Merfeld**'s sensorimotor comparator function, with its continuous feedback so as to minimise discrepancies between the goals and outputs of action, takes place prior to the time at which such action outputs come to be represented in conscious perception. The lag between the two processes is defined within the model (see the target article and Gray et al. 1997a) as being of the order of 100 msec. Continuous feedback so as to minimise discrepancies would be counterproductive in a device whose task precisely is to detect discrepancies, so that the threat that these pose can be properly evaluated and, potentially, responded to in a different manner.

For clarity of exposition, I have talked above as though there is a simple sequence in which a first phase of on-line sensorimotor integration is followed, ca. 100 msec later, by a phase of conscious perception. In fact, however, I see these two processes, of (unconscious) sensorimotor integration and (conscious) perception as both continuing in parallel (Milner & Goodale 1995). Furthermore, as stressed by McCrone (1999), the lateness of conscious perception can in many cases undergo significant compensation from anticipatory extrapolation along predictable trajectories in sensorimotor space. If the conscious perceptual process does not detect significant mismatch or threat, there is no need for interruption in the continuing sensorimotor process. However, if mismatch is detected, the conscious process is able to override ongoing sensorimotor programs and bring them to a halt.

A further point made by **Merfeld** is that direct evaluation of the state of the external world is impossible; rather, the only comparison possible is that between actual sensory signals and expected sensory signals. This point is undoubtedly correct. In a very real sense, the external world that we appear to perceive exists only inside our brains, being constructed on the basis of just those sensory signals and feedback from action which provide the inputs to the sensorimotor compartor function upon which Merfeld's commentary rests. There is no conflict between his and my views on this point. Put simply, I see the need for a further comparator function following upon Merfeld's, one that perhaps takes as its inputs the outputs from his. That function adds, for as yet mysterious reasons (the hard problem), a layer of conscious perception over a process of sensorimotor integration that gets on very nicely, thank you, without any conscious awareness at all (Milner & Goodale 1995).

References

Amit, D. J. (1995) The Hebbian paradigm reintegrated: Local reverberations as internal representations. *Behavioral and Brain Sciences* 18:617–57. [IV]

Black, I. (1994) *Information in the brain – a molecular perspective.* A Bradford Book. [WD]

Bridgeman, B., Van der Heijden, A. H. C. & Velichkovsky, B. M. (1994) A theory of visual stability across saccadic eye movements. *Behavioral and Brain Sciences* 17:247–92. [DMM]

Chalmers, D. (1996) *The conscious mind: In search of a fundamental theory.* Oxford University Press. [rJG]

DeLancey, C. (1996) Emotion and the function of consciousness. *Journal of Consciousness Studies* 3:492–99. [IV]

Dichgans, J., Held, R., Young, L. R. & Brandt, T. (1972) Moving visual scenes influence the apparent direction of gravity. *Science* 178:1217–19. [DMM]

Edelman, G. (1994) *Bright air, brilliant fire.* Penguin Books. [IV]

Flanagan, O. (1992) *Consciousness reconsidered.* MIT Press. [IV]

Freeman, W. J. (1995) *Societies of brains.* Erlbaum. [WD]

Frith, U. (1989) *Autism: Explaining the enigma.* Blackwell. [MHJ]

Gray, J. A. (1982a) *The neuropsychology of anxiety: An inquiry into the functions of the septo-hippocampal system.* Oxford University Press. [rJG]

(1982b) Précis of The neuropsychology of anxiety: An inquiry into the functions of the septo-hippocampal system. *Behavioral and Brain Sciences* 5:469–84. [rJG]

(1995) The contents of consciousness: A neuropsychological conjecture. *Behavioral and Brain Sciences* 18:659–722. [VI]

(1999) The hard question of consciousness: Information processing versus hard wiring. In: *Neuronal bases and psychological aspects of consciousness,* vol. 8, eds. C. Taddeo-Ferretto & C. Musio. World Scientific. [rJG]

Gray, J. A., Buhusi, C. V. & Schmajuk, N. (1997a) The transition from automatic to controlled processing. *Neural Networks* 10:1257–68. [rJG]

Gray, J. A. & McNaughton, N. (2000) *The neuropsychology of anxiety,* 2nd edition. Oxford University Press. [rJG]

Gray, J. A., Williams, S. C. R., Nunn, J. & Baron-Cohen, S. (1997b) Possible implications of synaesthesia for the hard question of consciousness. In: *Synaesthesia: Classic and contemporary readings,* eds. S. Baron-Cohen & J. E. Harrison. Blackwell. [rJG]

Grusser, O. J. (1994) Early concepts on efference copy and reafference. *Behavioral and Brain Sciences* 17:262–65. [DMM]

Hein, A. & Held, R. (1961) A neural model for labile sensorimotor coordination. *Biological Prototypes and Syntheticc Systems* 1:71–74. [DMM]

Held, R. (1961) Exposure history as a factor in maintaining stability of perception and coordination. *Journal of Nervous and Mental Disease* 132:26–32. [DMM]

Horn, S. & Munafò, M. (1997) *Pain: Theory, research and intervention.* Open University Press. [VI]

Humphrey, N. (1983) *Conscious regained: Chapters in the development of mind.* Oxford University Press. [VI, rJG]

(1993) *A history of the mind.* Vintage. [VI, rJG]

Jackendoff, R. (1987) *Consciousness and the computational mind.* MIT Press. [MHJ]

McCrone, J. (1999) *Going inside: A tour round a single moment of consciousness.* Faber and Faber. [rJG]

Merfeld, D. M. (1995a) Modeling human vestibular responses during eccentric rotation and off vertical axis rotation. *Acta Otolaryngologica Supplement* 520:354–59. [DMM]

(1995b) Modeling the vestibulo-ocular reflex of the squirrel monkey during eccentric rotation and roll tilt. *Experimental Brain Research* 106:123–34. [DMM]

Merfeld, D. M., Yong, L., Oman, C. & Shelhamer, M. (1993) A multi-dimensional model of the effect of gravity on the spatial orientation of the monkey. *Journal of Vestibular Research* 3:141–61. [DMM]

Milner, A. D. & Goodale, M. A. (1995) *The visual brain in action.* Oxford. [rJG]

Nagel, T. (1974) What is it like to be a bat? *Philosophical Review* 83:435–50. [VI]

Searle, J. R. (1980) Minds, brains, and programs. *Behavioral and Brain Sciences* 3:417–57. [WD]

Sperry, R. (1950) Neural basis of the spontaneous optokinetic response produced by vision inversion. *Journal of Comparative and Physiological Psychology* 43:482–9. [DMM]

Toates, F. (1995) Open peer commentary: On giving a more active and selective role to consciousness. *Behavioral and Brain Sciences* 18:700–701. [VI]

Varela, F., Thompson, E. & Rosch, E., eds. (1993) *The embodied mind.* MIT Press. [WD]

Velmans, M. (1991) Is human information processing conscious? *Behavioral and Brain Sciences* 14:651–726. [rJG]

Velmans, M. (1995) The limits of neurophysiological models of consciousness. *Behavioral and Brain Sciences* 18:702–703. [VI]

Von Holst, E & Mittelstaedt, H. (1950) Das Reafferenzprinzip (Wechselwirkungen zwischen Zentralnervensystem und Peripherie). *Naturwissenschaften* 37:464–76. (English translation: [1980] The reafference principle. In: *The organization of action,* ed. C. R. Gallistel. Wiley. [DMM]

Weiskrantz, L. (1997) *Consciousness loss and found.* Oxford University Press. [rJG]

https://doi.org/10.1017/S0140525X01213934 Published online by Cambridge University Press