Unity of Consciousness and Bi-Level Externalism*

BERNARD W. KOBES

In a conference hotel recently I accompanied a philosophical friend on a shopping expedition for some bold red and green wrapping papers, to be displayed as props in his upcoming talk on sensory qualia. Most examples used in philosophical discussions of consciousness have this static, snapshot character. The vehicle for a conscious perception of red wrapping paper may be a pattern of activation in a region of the brain devoted to visual input. But while walking, a person can turn her head, notice the wrapping paper, reach out, and grasp it. Our survival often depends on our ability to negotiate a changing environment by simultaneous streams of perception and action. What do vehicles of conscious content look like when there is realistically complex interaction with the world?

A dynamic view requires, according to Susan Hurley, a 'twisted rope', the strands of which are continuous multi-modal streams of inputs and outputs looping into the environment and back into the central nervous system. The twist in the rope is analogous to the unity of consciousness. The contents of conscious perceptions depend directly— non-instrumentally— on relations among inputs and outputs; similarly for conscious intentions.[1] We are evidently a long way from static red and green sensory qualia.

Hurley's book is a work of formidable ambition and multi-disciplinary erudition. She marshals wide-ranging literatures in the philosophy of mind, cognitive psychology, cognitive neuroscience, and dynamic systems theory. Her Kant and Wittgenstein scholarship is also impressive. Each chapter stands on

*Critical Notice of Susan L. Hurley, *Consciousness in Action*. Cambridge, MA: Harvard University Press, 1998. Pp. 528. ISBN 0-674-16420-2.

Part of this essay was presented in response to Susan Hurley's presentation, 'Perception, Dynamic Feedback, and Externalism ', at the American Philosophical Association, Eastern Division meeting, Washington D.C., 29 December 1998. I am grateful to Susan Hurley for her comments on that occasion. I also thank Ted Belledin and the Consciousness Discussion Group at Arizona State University.

Address for correspondence: Department of Philosophy, Box 872004, Arizona State University, Tempe AZ 85287-2004, USA.
Email: kobes@asu.edu

[1] In this paper I follow Hurley in using 'intention' only in its ordinary sense related to action, and not as a cognate of the technical word 'intentionality' for *aboutness*.

Mind & Language, Vol. 15 No. 5 November 2000, pp. 528–544.

its own, but the ideas interlock to create a large, unified vision. The whole is a dense fabric of argument, with recurring and criss-crossing themes. As a result it is sometimes difficult to find canonical formulations of her central concepts and arguments; moreover, the book takes a certain stamina to read through, and one often suspects that her argument could have been made more crisply without loss of force. But Hurley succeeds in bringing to the table a feast of new questions about the unity of consciousness (a major theme of Part I of her book), and the relation of perception and action to the dynamics of input and output (a major theme of Part II); and this must be acknowledged even by those who doubt, as I do, the adequacy of many of her particular answers to these questions.

The first of the four sections to follow describes briefly some ways in which Hurley's account is meant to make the self 'reappear out in the world' (p. 14). In the second section I criticize Hurley's account of the relation between normative constraints of coherence and the distinction between conscious unity and disunity. The third section suggests that perhaps the self has *already* 'reappeared out in the world' with the work of Putnam, Burge, and others who have used Twin Earth thought experiments to argue for externalism. Hurley rejects those thought experiments; I argue that she does so for mistaken reasons, and that Twin Earth Externalism would complement her other broadly externalist themes. In a final section I comment briefly on the theme of external feedback loops and the idea that vehicles of consciousness may loop outside the body.

1. The Reappearing Self

Hurley's views are broadly externalist, in that they characterize various aspects of mind and self in either a world-involving way or a norm-involving way, where worlds and norms are themselves independent of any purely internal subjective realm. A recurring motif is that the self is 'receding', 'hidden', or 'lurking' on standard, bad philosophical views, but that it 'reappears out in the open, embodied and embedded in the world' on her view (p. 3; cf. p. 14, pp. 335–337).[2]

Characteristic of standard bad views, and the chief whipping boy throughout the book, is the 'Input-Output Picture'. This is the tacit or explicit identification of perception with input, and of action with output, as in the phrase 'perceptual input and behavioral output'. Input and output, on Hurley's view, are concepts at the sub-personal level; they belong to a theoretically postulated system of causal processes that underlie the mental. Perception and action are concepts at the personal level; they belong to a norm-governed system of

[2] Another notable philosophical effort to see cognition as embodied and continuous with the world is Clark (1997). For a related externalist metaphysics of mind and self see Clark and Chalmers (1998). Neither of these, however, is explicitly about consciousness.

attributions to the whole person. The Input-Output Picture is an instance of an ingrained general tendency to confuse personal-level with sub-personal-level concepts. It also promotes, according to Hurley, a conception of the self as lurking hidden in a central zone, with perceptual input and behavioral output serving as buffers between self and world.

Hurley advocates replacing the Input-Output Picture with her Two-Level Interdependence View. On this view, conscious perceptual contents are a function of relations between inputs and outputs in a complex dynamic feedback system, and not of inputs alone. Conscious intentional contents are likewise a function (albeit a *different* function) of relations between inputs and outputs, and not of outputs alone. It should be possible in principle, then, to have cases in which inputs are held constant while outputs are varied, with the result that perceptual consciousness varies. Consider the paralyzed eye phenomenon: if a subject whose eye muscles are paralyzed tries to shift his gaze sideways, he will fail, and so there is no change in inputs to his perceptual system. Yet he may experience the world as shifting sideways (p. 344). Conversely, it should be possible in principle to have cases in which outputs are held constant while inputs are varied, with the result that conscious intentions vary.

Hurley constructs thought experiments inspired by observed experimental and clinical phenomena to argue that these surprising dependencies are indeed possible in principle (pp. 344–362). Of course action is typically instrumental to new perceptions, and perception is often instrumental to a change of action. But Hurley argues that her surprising dependencies can be 'brute'—non-instrumental—even though the cases presumably do involve feedback in the form of a motor signal or efferent copy internal to the brain.

The Two-Level Interdependence View broadens the sub-personal bases for personal-level perceptual and intentional contents, as compared to the Input-Output Picture. But its externalist flavor comes out most strongly when supplemented by Hurley's argument against the assumption that it is always possible to embed an actual and realistically active creature in a different environment while preserving the identity of its central processes. Nervous systems are 'knit . . . causally into their environments' (p. 328), and the causal loops whereby output feeds back to input are of various orbits, some extending far outside the skin. The sub-personal vehicles of consciousness may include external processes, looping out into the environment of the creature and back into its central nervous system. These themes are sub-personal analogues of the reappearing self (pp. 328–335, 371).

I will return to these issues, but first I want to focus on another important way in which Hurley has the self reappear 'out in the open'. Hurley argues in neo-Kantian fashion that the subjective perspective, as traditionally conceived, is not adequate for the theory of conscious unity. The argument concludes that objective factors are needed to make sense of the distinction between unified and disunified consciousness. Among such objective factors at the personal level are constraints of normative coherence. Normative constraints are objective in the sense that they do not derive from the character of subjective experience alone; they are in the public domain (p. 62). Unity and disunity of consciousness are, in that respect, essentially world-involving.

2. Co-consciousness and Normative Coherence

As I look out the window, I simultaneously see the green leaves of a tree and taste the warm tea I am sipping. These two conscious states are not only simultaneous and in the same human being; they are also, in an obvious and ordinary sense, 'together' or 'united' in the same conscious stream. We can imagine there being two simultaneous conscious states in the same human being that are not together in this way; perhaps this occurs in split-brain patients under certain experimental conditions. Hurley calls this relation of togetherness, 'co-consciousness'. (Derek Parfit (1984) uses the word in this sense, and the *Oxford English Dictionary* cites an early use by William James (1909, p. 268) of the word 'co-consciously', in the relevant sense.) But what is the nature of the co-consciousness relation?

Attributions of states and events at the personal level are, for Hurley, governed by normative constraints; moreover, the states and events themselves, at the personal level, are in part *constituted* by normative constraints. (In these views, Hurley is influenced by Donald Davidson.) The unity or disunity of consciousness is at least partly a personal-level phenomenon, and it too is at least partly constituted by normative constraints. Consistency, inferential integration, and means-end rationality are three kinds of normative constraint that help constitute the unity of consciousness (p. 124). In support of her view Hurley cites our practices in interpreting split-brain patients. When their behaviors appear to manifest mutually incoherent mental states, as might occur with simultaneous left hand and speech behaviors under experimental conditions, we interpret those mental states as non-co-conscious (p. 114).

Of course it is possible to imagine two creatures, A and B, each of whom is normatively coherent, and whose conscious states are exactly alike. In that case A's and B's conscious states will be perfectly coherent with respect to each other, but that will not in the least tend to make A's states co-conscious with B's. Hurley's view, then, is that normative coherence is a necessary condition for unity, but not a sufficient condition. Normative constraints are one-directional (p. 120). (They are also defeasible; in unusual cases, normatively incoherent states might nevertheless be co-conscious (p. 115). But Hurley usually allows this qualification to remain tacit, and I will follow her in this.)

Now it is emphatically not Hurley's view that normative incoherence is merely evidence for disunity. Normative constraints are constitutive, not merely evidential. Norms do not merely constrain attributions of unity and disunity; they constrain the very nature of unity and disunity. Hurley says, of the constraint forbidding inconsistent content in one consciousness at a time,

> Its role is demonstrated by the way in which disunity is attributed in various neuropsychological cases. But its role is not merely one of revealing disunity. It is not just a reliable generalization that inconsistent contents are separate. Where a normative constraint does apply, it has a constitutive role: it makes bad sense in principle to suppose that such inconsistent contents could be co-conscious. (p. 115)

Normative constraints are 'a priori' (p. 115), according to Hurley; '[e]xtreme enough normative incoherence . . . can *underwrite* disunity', and 'incoherence *makes for* disunity' (p. 120, my emphasis). So her position is that normative coherence of conscious states is a constitutive but defeasible necessary condition for unity of consciousness. Equivalently, normative incoherence of conscious states is a constitutive but defeasible sufficient condition for disunity of consciousness.

Hurley further suggests that the notion of a *dynamic singularity* at the sub-personal level might complement normative constraints at the personal level. Dynamic singularities are 'structural singularities in the field of causal flows characterized through time by a tangle of multiple feedback loops of varying orbits' (p. 2). The pattern of motor outputs and different types of sensory feedback 'looks like a tangle or knot, centered on the organism and moving with it: a singularity in the field of causal flows' (p. 307). This notion of a dynamic singularity derives from the technical theory of complex dynamic systems, but it is never made fully clear whether Hurley intends the notion literally as it appears in that technical theory, or whether it is meant as a picture or metaphor drawn from the technical theory. It is intended to be a functional notion with a neuroanatomical realization in terrestrial animals, but there need be no corresponding neuroanatomical unity, such as a particular structure or location in the brain (pp. 193–194).

The proposal, then, is that unity of consciousness is a multi-level phenomenon: among conscious states, normative coherence at the level of content, and sameness of dynamic singularity at the sub-personal level, are individually necessary and jointly sufficient for their unity (p. 216; p. 130 n. 22). Normally these two factors work together, but in unusual cases they may pull apart. In a case of complete duplication of contents in two subjects, the two dynamic singularities constitute disunity despite perfect normative coherence. In a case of multiple personality, there may be only one dynamic singularity, but the normative incoherence may constitute disunity of consciousness (p. 215).

Now there is a very ordinary observation about co-consciousness that appears to create a difficulty for Hurley's theory. It must be a common experience that when two or more of one's beliefs that are mutually inconsistent are brought together in consciousness, that is, when the beliefs *become co-conscious*, a process of rational integration occurs as a result: the beliefs' contents change, or one's degree of belief in one or both of the contents is diminished. It is plausible that the beliefs' becoming co-conscious *allows* certain integrative processes to work on them, and that these integrative processes tend to raise overall normative coherence. In fact, the beliefs' becoming co-conscious is plausibly antecedent, in temporal order and in the order of explanation, to their normative integration. (This is so even if it was the beliefs' mutual inconsistency that caused them to become co-conscious in the first place.) But then it is difficult to see how the norms that govern processes of integration could *constitute*, even in part as Hurley's theory would have it, the co-consciousness that allows those integrative processes to get started.

These observations suggest a quite different framework for the relation between normative constraints and co-consciousness, one that Hurley seems not to consider in her book. It is still true that co-consciousness can be characterized as that species of 'togetherness' of conscious states that permits the operation on them of such norm-governed integrative processes as are available to the subject. Such a characterization may state a theoretically important truth about co-consciousness; for example, we could inquire into whether it states a *biological function* of co-consciousness. But from a semantic standpoint, the characterization at best fixes the reference of the expression 'co-conscious'. It appears to leave entirely open the question of the *nature* of the co-consciousness relation.

This alternative framework is supported by a further observation: in an entirely ordinary sense, there is *no barrier at all* to incoherent contents being co-conscious. At this moment, for example, the mutually incoherent conscious propositional contents 'Snow is wholly and completely white' and 'Snow is wholly and completely orange' are co-conscious in my mind; I can entertain these contents consciously and simultaneously without strain. So there is an ordinary sense in which the normative incoherence of conscious contents *as such* is not at all a sufficient condition for disunity (not even a defeasible sufficient condition); and *a fortiori* normative incoherence of contents as such is not a *constitutively* sufficient condition for disunity.

Examples of this sort can be found in other modalities of mental representation. They are easily found for mental imagery, but let us proceed directly to what may seem the most difficult kind of case for my purposes, that of current conscious perception. I focus my gaze on a distant doorknob and hold up my pen directly in the line of sight, so that I see the pen 'double'. One way to interpret this example is that I have a conscious visual perception *as of holding exactly two pens*. But this is normatively incoherent with my co-conscious tactile perception *as of holding exactly one pen*. A different way to interpret this example is that I have a conscious visual perception *as of holding exactly one pen and it is just to the left of the doorknob (relative to my line of sight)*; but on this interpretation I have a co-conscious visual perception *as of holding exactly one pen and it is just to the right of the doorknob (relative to my line of sight)*. On either interpretation of the example, I have co-conscious normatively incoherent perceptual contents. (So I find unconvincing Hurley's claim that 'there can be no experience with the internally incoherent content: I am seeing just one light and it is wholly red and I am seeing just one light and it is wholly green' (p. 118)). Again, normative incoherence of conscious perceptual contents is not a sufficient condition for disunity, and *a fortiori* it is not a *constitutively* sufficient condition for disunity.

At this point it may be objected in defense of Hurley's theory that my examples are all ones in which the subject does not and indeed cannot *believe* the normatively incoherent conscious contents (except perhaps in pathological cases in which the defeasibility condition kicks in). Moreover, the objector may continue, whether or not conscious contents are normatively coherent is sensitive to whether or not they are believed; if they are merely *entertained* or *experienced*, then they simply do not count as normatively incoherent. Once normative incoherence is understood in this way, we can say that normative incoherence of conscious contents is a constitutively sufficient condition for disunity.

There are three responses to this objection. (I will cast the responses with an eye to Hurley's normative constraint of consistency, but I believe they can be generalized to the normative constraints of inferential integration and means-end rationality.) The first and simplest response is that it is difficult to see how normative constraints that apply only to a proper subclass of a relation can be held to constitute, even partly, the wider relation. Mutually inconsistent conscious desires, for example, can be co-conscious; Hamlet desires *to be*, and also desires *not to be*, and these inconsistent desires may easily be co-conscious. It is difficult to see how the co-consciousness relation that holds between Hamlet's desires can be constituted, even partly, by normative constraints that apply only to beliefs. Hurley's normative constraints may be cast in conditional form (*if* x and y are beliefs, *then* . . .), and may thus be said to apply to all conscious states. But the essential point would remain: it is difficult to see how normative constraints that apply *non-trivially* only to a proper subclass of a relation can be constitutive, even partly, of the wider relation.

The second response begins by noting that the following two conscious propositional contents can easily occur simultaneously in a single conscious stream: 'I believe that snow is wholly and completely white' and 'I believe that snow is wholly and completely orange'—provided these conscious contents are not themselves believed. Belief, we may assume for present purposes, is a relation that holds between persons and conscious contents. It would be a mistake to treat the belief relation as part of the conscious content believed; that would be to confuse the relation with one of its relata, and it would open up the question of whether or not the augmented content is believed. The belief relation to a conscious content is not a further bit of conscious content. Instead, it is a normatively constrained functional relation to conscious content. Since Hurley's normative constraint of consistency is more appropriately cast as a constraint on what can be co-consciously *believed*, and since the belief relation is not itself among the contents of consciousness, the norm does not constrain what contents can be co-conscious, and hence does not constitute the co-consciousness relation.

The third response is that, having made explicit that it is only via the handle of the belief relation that the norm of consistency gets any sort of grip on co-consciousness, we should now ask whether Hurley gets the direction of explanation right. We may agree that, normally, 'the most basic connection between unity and coherence is one-directional: some minimum of coherence may be necessary for unity . . ., but it is not sufficient for unity' (p. 120), if this means only that we have a conditional and not its converse. But given Hurley's doctrine of the constitutive bearing of normative constraints on questions of unity, this gets cast in either of two ways: 'normative coherence is constitutively necessary for unity', and 'normative incoherence is constitutively sufficient for disunity', and both expressions line up the direction of explanation *from* coherence/incoherence *to* unity/disunity.

Now the point is that the reverse direction of explanation is more plausible: normally, 'unity is sufficient for, and helps explain, normative coherence', and 'disunity is necessary for, and helps explain, normative incoherence'. This direction is more plausible because it is easy to see how the explanation works, in the case of the norm of consistency, and why the belief relation should play a special role: when inconsistent contents become co-conscious, available processes of integration are able to work on the conscious states, normally causing the subject's degree of belief in one or both of the contents to be diminished. No comparably clear explanation of the special role of the belief relation is available under Hurley's direction of explanation.

This normal explanatory connection is, of course, defeasible. It may happen, in abnormal cases, that inconsistent but co-conscious contents are not integrated into a coherent whole. I believe that this point is more apt to help us understand actual cases of multiple personality disorder than Hurley's suggestion about multiple personalities. Her suggestion, again, is that multiples might be cases in which the dynamic singularity criterion is satisfied, but in which the failure of the normative coherence criterion constitutes *failure of unity* across alters (pp. 121–122, p. 215). Since unity is co-consciousness writ large (pp. 40–41, 88–89, 102–106), presumably she has in mind simultaneous conscious states that are not co-conscious across alters. Perhaps Hurley intends this suggestion only as a thought experiment to illustrate her views on unity. But it is worth pointing out that clear evidence of simultaneously conscious but non-co-conscious states in multiples is difficult to come by. Most multiples display haziness about the past, 'missing time', or amnesia, which suggests a model in which the distinct personalities are only successively rather than simultaneously conscious. Some multiples, on the other hand, display what is called, in the literature on multiple personality disorder, 'co-consciousness' (n.b.!). Ian Hacking (1995, p. 27) illustrates:

> The alters argue with each other, snarl, or console. One alter may be

> out and yet have another alter yammering away beside the left ear, telling her what a ninny she is. Many therapists try to introduce different alters to each other, believing that thoroughgoing co-consciousness is a necessary step toward integration.

When alters are co-conscious in this sense, then conscious states across those alters are co-conscious in the sense of Parfit and Hurley, despite being normatively incoherent. (Our notion of co-consciousness—being together in the same conscious stream—is more basic than the distinction between attributing an experience or mental act to oneself or to another. A schizophrenic's conscious experience, as of a voice which he attributes to another, is co-conscious with the rest of his conscious experience.) Co-consciousness is the same relation in normal cases as it is in pathological cases in which normatively constrained integrative processes fail.

I have argued that co-consciousness is prior to the integrative processes that it normally permits, and that it is constituted independently of the normative constraints that govern those processes and conscious contents themselves. Perhaps these conclusions are no surprise. An externalism based on norms of coherence seems apt for contents and integrative processes. It should seem less apt for basic structural aspects of consciousness, such as co-consciousness.

Let me end this section by briefly noting that Hurley also argues for a different way in which broadly normative considerations may bear on unity of consciousness. She imagines two subjects who both lack a functioning corpus callosum, and who both use cross-cueing behavior and access movements to get information from one hemisphere to the other. Let them use the same external methods, and to the same degree. Now one subject is a recent commissurotomy patient, while the other has callosal agenesis, a condition in which the corpus callosum has been absent since birth. The argument suggests that cross-cueing and access movements might have a different status in these two patients. The acallosal may count as unified, and the commissurotomy patient not, purely in virtue of differences in their distant histories of using cross-cueing and access movements (pp. 189–193).

Limitations of space prevent me from saying much about this interesting argument, other than to point out that it seems to follow from its conclusion, together with Hurley's two-level account of unity, that distant historical considerations must enter into the individuation of dynamic singularities at the present time. I do not know if Hurley would accept this apparent implication. In any case, the result strikes me as troubling, given that the notion of a dynamic singularity is supposed to derive from the technical theory of complex dynamic systems.

3. Twin Earth and Bi-level Externalism

We have seen that Hurley's view of mind and self has a broadly externalist character at both the personal and sub-personal levels. At the personal level externalism is most clearly manifested in her account of the unity of consciousness. Though at the personal level, the account directly concerns the structure, rather than the contents, of consciousness. At the sub-personal level externalism is most clearly manifested in Two-Level Interdependence, supplemented by an externalist account of input-output relations and vehicles of consciousness. But none of this quite amounts to a full-fledged externalism with respect to conscious contents. So one might have expected Hurley to supplement her various externalist views with further ones based on Twin Earth thought experiments, which have been much used in arguments purporting to show that mental contents are typically directly fixed in part by relations that the subject bears to his natural and social environments. Such a full-fledged Bi-level Externalism would seem to be in the spirit of Hurley's efforts to bring the self out into the open.

Instead, however, we find Hurley mounting a sustained attack on the use of Twin Earth thought experiments in debates between internalists and externalists. On her account of the dialectical situation, both internalists and externalists use Twin Earth thought experiments to elicit intuitions in support of their respective views of the determinants of mental content. But in so doing, both camps tacitly make the Duplication Assumption. This is the idea that it is always possible, in principle, to imagine a subject in a different environment from the one that he actually inhabits, certain features of the environment being systematically inverted, while holding fixed all of the subject's physical states internal to some significant boundary, such as the skin, or the central nervous system (pp. 245–249, 294–298).

Hurley argues that the Duplication Assumption is false, or at least highly dubious. She does this by setting up a series of increasingly complex Twin Earth cases in which the imagined twin is fitted with peripheral technologies that preserve the identity of his internal states with those of the Earthly subject, despite environmental inversions. In Ned Block's well-known case of Inverted Earth, for example, the normal colors of external things are inverted (the sky is green, grass is blue, etc.), but the twin subject is fitted with retinal implants that change the frequencies of incoming light to what they would be for a normal subject on Earth. Thus the subject and his twin remain internally identical despite the environmental inversion. In Hurley's 'Mirror World', in which every physical thing in the subject's environment is left-right inverted (reflected through the subject's sagittal plane), the subject is fitted with a number of such inverting technologies (visual, motor, proprioceptive, etc.) so as to maintain subjective identity with his normal counterpart on Earth (pp. 256–260, 299–314).

Finally, in Hurley's 'El Greco' worlds, all physical objects are vertically elongated; their heights are doubled as compared to the heights of their Earthly counterparts. Hurley argues that if the El Greco subject, Twin Dom, is to be internally identical to his Earthly counterpart Dom, but also well-adapted to the El Greco environment, then the inverting technologies implanted at the periphery of his sensory and motor systems would have to compute complex relations among multi-modal motor-to-sensory feedback channels in order to select which adjustments to make. But such computations are, in the relevant functional sense, 'central' in nature, despite occurring at the periphery of the nervous system. So we have not really succeeded in holding fixed central cognitive processes in Dom and Twin Dom, as the Twin Earth methodology would require (pp. 314–325). This is the core of Hurley's argument against the Duplication Assumption; I will return to it.

One salutary lesson to take from Hurley's discussion is that it behooves Twin Earth builders to construct with great care their twin environments and the subjects who have to live in them. It is a serious question whether the details can be coherently completed in Burge's (1986a, 1986b) cracks and shadows case, for example, or in Davies's (1993, 1996) case in which ellipses give rise to the same internal states that circles give rise to on earth– a case that somewhat resembles Hurley's El Greco case.

Hurley, however, wants to establish a more ambitious critical outcome: in casting doubt on the Duplication Assumption, she takes herself to have undercut the standard debate between internalists and externalists, who are both committed to it. 'It is the shared Duplication Assumption and its presupposed boundary that *gives substance* to the disagreement about whether the dependence of content on the external world is direct– brutely relational– or merely instrumental.' (p. 297, my emphasis) Presumably, if the Duplication Assumption turns out to be false, then externalism as it appears in the conclusions of Twin Earth arguments—Twin Earth Externalism—would be somehow defective. But it is not entirely clear exactly what she thinks that defect would be. Would Twin Earth Externalism be less than fully intelligible in virtue of resting—along with its opposite—on a failed presupposition? Or would it suffer from some other defect, such as being unestablished (even if true), or uninteresting (even if true)?

One source of this interpretive difficulty is Hurley's tendency to lump together Twin Earth arguments indiscriminately. There are by now a variety of Twin Earth thought experiments that have been put forward to show that mental contents of one type or another are in part externally fixed. Some involve concepts (e.g., 'aluminum') for natural kinds in the environment; some involve concepts (e.g., 'arthritis') whose extensions are fixed by expert usage in the linguistic environment; some involve color and spatial perception. These thought experiments do not necessarily stand or fall together. There may be interesting differences among kinds of mental content with respect to whether and how they are externally fixed.

The Duplication Assumption evidently has a strong quantificational structure: 'The duplicationist assumes that it is *always* possible in principle to duplicate internal physical processes within some philosophically significant boundary in a distorted environment' (p. 323, my emphasis; cf. pp. 297, 302, 324, 325, and 328). A bit more explicitly, the assumption seems to be that, given any subject S adapted to an earthly environment, and given any Twin Earth that is symmetrically inverted in some respect relative to the earthly environment, it will be possible to imagine Twin-S, a duplicate of S's body inside some significant boundary, functioning successfully and naturally on Twin Earth, perhaps with the aid of input and output adjusters.

Hurley is plainly mistaken in supposing that Twin Earth thought experiments depend on the Duplication Assumption. For one thing, the possibility of duplication in the case at hand is often argued for, not simply assumed. But the main point is that *no* Twin Earth thought experiment assumes or entails that duplication is possible for *every* possible pair of symmetrically inverted environments. Each thought experiment must be evaluated on its own terms. The required duplications do not necessarily stand or fall together. Some may succeed, others fail. It could turn out that, for some categories of mental content, duplication cannot be achieved, while for others, it can. This would in itself be a philosophically significant result, raising interesting questions for further investigation.

The Duplication Assumption includes the assumption of a 'significant boundary' for fixing content, which Hurley argues against (pp. 282–283, 295–297, 323–325, 336). The Twin Earth methodology, then, is saddled with the assumption of such a significant boundary, either at the skin or at the central nervous system. Again, I think this is a mistake. Suppose an externalist presents a Twin Earth thought experiment to argue for externalism. Of course this involves imagining duplication within the skin, or within the central nervous system. But at no point does the argument assume that this is a significant boundary. In fact, the conclusion of the argument is precisely that the boundary does *not* have a certain significance—that of fully determining a particular kind of mental content.

Suppose it is replied that, although no single Twin Earth thought experiment makes Hurley's Duplication Assumption, the maximally ambitious Twin Earth Externalist claim that *any* mental content can be shown by Twin Earth thought experiment to be in part externally fixed *does* make the Duplication Assumption. Again, this would be a mistake. The maximally ambitious Twin Earth Externalist is only committed to the following: for any subject S in an Earthly environment, and for any one of S's mental contents, there exists *some* possible twin environment in which a replica of S's body or central nervous system can function naturally, such that Twin-S intuitively lacks the relevant mental content. This is a much weaker commitment, in virtue of its quantificational structure, than Hurley's Duplication Assumption, which requires duplication for *any* pair of symmetrically inverted environments.

For example, suppose it were decisively shown that duplication is not possible in any of Hurley's El Greco worlds. Would maximally ambitious Twin Earth Externalism be thereby refuted? No. First we would need to identify the relevant mental contents; suppose they are visual perceptions of shapes and distances. Can the contents of visual shape and distance perceptions be shown to be externally fixed by Twin Earth thought experiments? The burden of proof would no doubt fall on the maximally ambitious Twin Earth Externalist, but the result that duplication is impossible in El Greco worlds would be, at best, only a start in showing that the burden cannot be met. To show that the maximally ambitious Twin Earth Externalist is bound to fail in the case of visual shape and distance perceptions, one would need some argument to extend the failure of duplication in the El Greco worlds to the failure of duplication in *any* twin world that tinkers with visual shape and distance perception. I myself cannot see, and there is nothing in Hurley's book to suggest, how such an argument might go.

Suppose, however, that Hurley did succeed in showing that no duplication relevant to external fixing of visual shape and distance perception is possible. There would still be a fallback externalist position available, on which it would still be true that *all* mental contents are in part externally fixed. According to the fallback position, we need a mix of externalist arguments. Twin Earth arguments would work for some (or most, or almost all) kinds of mental content. Other kinds of externalist arguments would be needed for the remainder. Less-than-maximally ambitious Twin Earth Externalism can be a key component of a larger, maximally ambitious externalist strategy.

This strategy can be elaborated. As Hurley rightly notes, local supervenience of a kind of content is not by itself a victory for internalism with respect to that kind of content (p. 297, n. 9). Properties instantiated in the environment might be brutely, not merely instrumentally, relevant to the instantiation of that type of mental content, in virtue of their explanatory priority, even if a certain type of internal physical organization necessarily accompanies it. The relevant mental content might supervene on internal states, but the best explanation of why this is so would take an explanatory detour: given those internal states, certain environmental states must necessarily obtain, and these fix the relevant mental contents directly. In this way, different kinds of externalist arguments about mental content can co-exist, and even complement each other in a divide-and-conquer strategy, or reinforce each other by consilience.

Hurley acknowledges that duplication of the sort required for a Twin Earth argument seems possible in Ned Block's Inverted Earth case, and also in her own case of Mirror World. Curiously, however, Hurley rejects less-than-maximally-ambitious Twin Earth approaches to externalism. She anticipates the reply (p. 329):

> So long as the issue arises and goes his way for some cases, he [the Twin-Earth Externalist] will be vindicated, even if it doesn't arise for other cases.

In response, she writes:

> But this reply doesn't go deep enough. If duplication is in principle

> problematic in certain types of case that are basic to what it is to have a perspective, we may begin to doubt the general significance of the brand of Externalism that assumes duplication even in other types of case where it is unproblematic.

This is cautiously phrased; nevertheless, the doubt expressed seems wholly gratuitous. Ignore for a moment the problem of extrapolating from the El Greco worlds to all twin cases relevant to certain contents 'basic to what it is to have a perspective', and suppose that twin cases cannot establish externalism for those contents. Why should that be taken to cast doubt on the 'general significance' of Twin Earth Externalism? We are given no reason to fear that the failure of twin cases for those perspectivally basic contents will spread to other contents. Nor are we given any reason to think that the success of twin cases for other contents would be a boring secondary phenomenon riding piggyback on more exciting varieties of externalism.

I take two morals from these observations. The first is that Hurley's discussion of the El Greco worlds does no clear damage to even the most ambitious Twin Earth Externalist program, and is very far from doing any damage to less-than-maximally ambitious Twin Earth Externalist programs. The second is that if different types of externalist argument can co-exist and even complement and reinforce each other with respect to different types of mental content, then so should different types of externalist argument with respect to mental contents and their sub-personal vehicles. Therefore, and in light of Hurley's deconstruction of the significance of skin or central nervous system boundaries at the sub-personal level, I urge Bi-level Externalism as a powerful and live option for philosophy of mind.

4. El Greco Worlds and External Vehicles

Hurley appears to take it as a constraint on the relevant kind of duplication that both the subject and his twin function normally, naturally, and successfully in their respective environments. A twin's success in negotiating his environment should not be due to a series of divine interventions or lucky accidents. Nor can we admit twins who are brains in vats, or who wear extensive virtual reality gear. Hurley excludes such cases from her discussion of the Duplication Assumption, and rightly so, since they would trivialize it.

I grant Hurley's point that the computations of input-output adjustments that would be needed in El Greco worlds relevantly resemble central mental processes. But Hurley assumes, implausibly, that they should be counted among *Twin Dom's* central processes. On her view, duplication fails in El Greco cases because the central mental processes of the (putative) twins would not be type identical, given the *extra* central processes that Twin Dom would enjoy in virtue of his input-output adjusters. But the topology of Twin Dom's feedback loops seems incompatible with ascribing those mind-like multi-modal processes to Twin Dom himself or to his mind, for those processes would not be *centred* (in the sense of a dynamic singularity, if you like) on the relevant unified subject, but rather on some point along an input or output channel. (If the transformations were performed by conscious homunculi, the homunculi would not be co-conscious with Twin Dom!) This is a point about the functional topology of the input-output adjustment mechanisms; the point would hold even if those devices happened to be physically located inside Twin Dom's central nervous system.

There is in any case a different and clearer route to rejecting duplication in El Greco worlds. Twin worlds need to run in parallel over a substantial stretch of time, during which the twins interact with their respective worlds normally, naturally, and successfully. Since the El Greco worlds are to be realistically complex, the differences between the El Greco world and the actual world will quickly multiply and spread. As the twins act on their worlds in normal ways, differences between the worlds will ramify in ways that the input-output adjusters cannot always anticipate. To ensure that Twin Dom's central nervous system continues to duplicate Dom's, the complex multimodal adjusters of input and output will have to amount to full-fledged virtual reality gear, which will give Twin Dom the *illusion* of inhabiting a world whose objects and events parallel those on Earth. But it will be a virtual reality that increasingly diverges from ordinary El Greco reality. It is not that the input-output adjusters invade Twin Dom's central processes, but that the adjusters must create for Twin Dom a 'notional world'.[3] Thus duplication of the relevant sort fails in the El Greco worlds.

The El Greco cases ingeniously and vividly illustrate Hurley's vision of complex dynamic feedback loops of varying orbits that 'knit nervous systems causally into their environments' (p. 328). This is what gives Two-Level Interdependence its externalist flavor. The strands of the 'twisted rope' are content-specific horizontal mental modules that include input and output as well as feedback by way of efferent copy, proprioception, and worldly re-afference (pp. 183–184, 406–412). The very vehicles of consciousness may loop outside the body, Hurley suggests, in her clearest and most explicit expression of Externalism at the sub-personal level (pp. 328–335, 371).

Hurley understands vehicles as 'differentially token-explanatory': the vehicle for a conscious content token *c* is what would be left of the full explanation of token *c* after we subtract out the more general explanation of why conscious contents of that *type* occur (pp. 28, 330). Suppose we accept an account along these lines. Then, I suggest, the notion of a *vehicle*, and in particular the notion of an *external vehicle*, is tied to the pragmatics of explanation. For explanation is notoriously multi-faceted; we can distinguish causal from

[3] The term 'notional world' is borrowed from Dennett (1982). For an externalist treatment of notional worlds, see Kobes (1990).

constitutive explanation, and proximal from distal causes. Perhaps the only interesting explanations are those that make a highly selective cut from among a welter of constituents or causal antecedents. Just where the cut ought to be made is a function of the pragmatics of explanation. Various notions of *vehicle* may emerge, depending on the theorist's explanatory motivations.

Suppose that a person S, while walking, turns her head, notices some bold red wrapping paper, reaches out, and grasps it; let *r* be her conscious visual perception as of red paper during that time. What is the *vehicle* for the content token *r*? Hurley's account directs us to start with a full explanation of the token *r* and subtract out the explanation of why conscious contents of that type occur. But we can imagine (at least) two kinds of explanation of the token *r*. (1) The explanation could take for granted all physical states and processes of the world, and aim at an account of *r* against that background—why is there the conscious token *r* in a world thus physically structured, as opposed, for example, to S's being a zombie (in the philosopher's sense)? (2) The explanation could aim at an account of how *r* is brought about and how it is sustained through the relevant time interval—why *r*, rather than the distinct contents *s* or *t*, given certain causal antecedents?

Relative to explanatory motivation (1), the vehicle for *r* may lie wholly within S's central nervous system. Relative to explanatory motivation (2), the vehicle for *r* may orbit beyond S's skin and back. It may have the character of a feedback loop that includes the red wrapping paper; any wholly internal putative vehicle may now seem unnaturally jagged and truncated. This relativization of the notion *vehicle* to explanatory motivation is inspired by Hurley's differentially token-explanatory account; it aims to be compatible with her chief claims and arguments, and it hopes to do better justice to both the unique contribution of the central nervous system and the dynamic, embedded character of consciousness.

This conjecture illustrates my view that Hurley's book is valuable not for answering old questions about consciousness but for revealing a landscape of new questions for many disciplines. She has illumined the co-consciousness relation, and softened boundaries that seemed to separate us perceivers and agents from our world—but these impressionistic phrases, and this critical review, can only hint at the sweep of her philosophical vision.

Department of Philosophy
Arizona State University

References

Burge, T. 1986a: Cartesian error and the objectivity of perception. In P. Pettit and J. McDowell (eds.), *Subject, Thought, and Context*. Oxford University Press, 117–136.

Burge, T. 1986b: Individualism and psychology. *The Philosophical Review*, 95, 3–45.

Clark, A. 1997: *Being There: Putting Brain, Body, and World Together Again*. The MIT Press.

Clark, A. and Chalmers, D.J. 1998: The extended mind. *Analysis*, 58, 10–23.

Davies, M. 1993: Aims and claims of externalist arguments. In E. Villanueva (ed.), *Philosophical Issues 4: Naturalism and Normativity*. Ridgeview Publishing Co., 227–249.

Davies, M. 1996: Externalism and experience. In A. Clark, J. Ezquerro, and J.M. Larrazabal (eds.), *Philosophy and Cognitive Science: Categories, Consciousness, and Reasoning*. Kluwer Academic Publishers, 1–33. (Reprinted in N. Block, O. Flanagan, and G. Güzeldere (eds.), *The Nature of Consciousness: Philosophical Debates*. The MIT Press, 1997, 309–327.)

Dennett, D. 1982: Beyond belief. In A. Woodfield (ed.), *Thought and Object*. Oxford University Press, 1–95. (Reprinted in D. Dennett, *The Intentional Stance*. The MIT Press, 1987, 117–202.)

Hacking, I. 1995: *Rewriting the Soul: Multiple Personality and the Sciences of Memory*. Princeton University Press.

James, W. 1909: *A Pluralistic Universe*. Longmans, Green, and Co.

Kobes, B.W. 1990: Individualism and artificial intelligence. In J.E. Tomberlin (ed.), *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*. Ridgeview Publishing Co., 429–459.

Parfit, D. 1984: *Reasons and Persons*. Oxford University Press.