

Entangled externalisms

Mark Sprevak
University of Edinburgh

Jesper Kallestrup
University of Edinburgh

10 January 2014

The debate between internalists and externalists is multifaceted, straddling vexed issues in contemporary philosophy. This chapter focuses on the distinction between content and vehicle as it pertains to the internalism/externalism debate in philosophy of mind and cognitive science. Whereas *content* internalism/externalism seeks to give an account of what makes mental states have the contents they have rather than some other contents or no contents at all, *vehicle* internalism/externalism seeks to give an account of the processes or mechanisms that enable mental states with contents to play a causal role in, for example, guiding behavior.¹ In general, we understand externalism as the negation of internalism.

What then is content internalism (CI)? The basic idea of CI is that the contents of mental states are narrow in the sense of supervening on internal features of individuals who are in those states. By ‘individual’ we henceforth understand a cognitive system, capable of being in content-bearing mental states. In the following, we are only concerned with the contents of beliefs and other propositional attitudes. We also assume throughout that physicalism is true of our world such that those internal features are physical. We later try to make precise the relevant notion of supervenience. In contrast, content externalism (CE) is the view that the contents of mental states are wide in that they fail to supervene on internal physical features of individuals.

CE is typically motivated by Twin-Earth-style cases. Suppose I utter the sentence ‘that apple is wholesome’ while pointing at the apple—call it apple₁—in front of

1. The terminology owes much to Hurley (2010).

me. What I have said is true iff apple₁ is wholesome. Now suppose an internal physical duplicate of me also utters the sentence ‘that apple is wholesome’ while pointing at a distinct yet superficially indistinguishable apple—call it apple₂—in front of him. What my duplicate said is true iff apple₂ is wholesome. Given that the truth-conditional contents of utterances of the same sentence by internal physical duplicates differ, those contents fail to supervene on internal physical features. On the assumption that the contents of beliefs are given by the sentences (and accompanying demonstrative identifications) used to correctly report those beliefs, then the corresponding mental contents also fail to supervene on such features. In response, proponents of CI might try to factor out a narrow component, shared by internal physical duplicates. Perhaps what my duplicate and I have in common is the belief that the demonstratively identified apple is wholesome, where the description ‘the demonstratively identified apple’ picks out different apples in different contexts of utterance. The dispute between CE and CI is not merely over the content of sentences containing demonstrative expressions. Friends of CE also hold that the content of sentences containing natural kind, or artifactual, terms fails to supervene on internal physical features. Instead, such content is wide in virtue of being partially individuated by environmental features to do with the microstructure of the relevant natural kinds or sociolinguistic facts about language use and speakers’ deferential dispositions. The same is supposedly true of mental content. Friends of CI have devised various strategies for resisting both claims, which will not detain us here.²

How about vehicle internalism (VI)? The vehicle of content is the physical item that has, or expresses, that content—for example, a sentence, if we talk about linguistic content, or some piece of cognitive architecture, if we talk about mental content. The basic idea of VI is that an individual’s mental processing is brain- or at least body-bound; cognitive processes and mental states are located inside the skin and skull of individuals. One can get an intuitive grip on VI by thinking of the mind, roughly speaking, as a sensation-cognition-action sandwich.³ Cognition is the ‘filling’ of this sandwich: cognition takes place *after* sensory input and *before* motor output. Since sensation and motor activity occur at bodily interfaces and cognition occurs between sensation and motor activity, it appears that cognition must occur inside the body or, rather, the physical processes that correspond to cognition must lie inside the body. In contrast, vehicle externalism (VE) claims that human cognition is neither brain- nor body-bound: our cognitive processes ‘extend’ outside the human body to include objects and processes in the external environment.

VE is motivated by a range of arguments. One argument, the extended-functionalism argument, begins with the widely held claim that functional structure is the essential feature of cognition. What makes a physical process a cognitive

2. For more details, see Kallestrup (2011).

3. Hurley (1998) calls this the ‘Input–Output Picture’.

process—say, of deductive reasoning, inductive reasoning or word association—is its informational states and the way in which those states are manipulated in the process. If a system has a mechanism with the right informational states and the right functional structure, that system counts as having the relevant cognitive process. This holds no matter what the states are made out of or where they occur. Extended functionalists argue that the requirements for cognition to occur are met, not only by human neural activity (as functionalists have claimed since the 1960s), but also by the conjunction of that neural activity and the use of external resources.⁴

To fix ideas, consider the well-trodden example of Otto. Imagine that Otto has a mild form of Alzheimer's and he always carries a notebook with him. When Otto needs to store new information, he always writes it down in his notebook, and when he needs to recall information, he always looks it up in his notebook. Advocates of extended functionalism argue that Otto's notebook, if used in a sufficiently reliable way, plays the same functional role in Otto's mental life as neural memory does for Otto's healthy counterpart, Inga. In Inga's case, the functional requirements of memory are fulfilled by her brain activity alone; in Otto's case, those requirements are fulfilled by the joint operation of Otto's brain, body and notebook. Otto's storage and recall of information from his notebook is, by functionalist lights, a case of extended cognition. In response, proponents of VI object that there are functional differences between Otto and Inga that show that Otto and his notebook do not fulfill the functionalist requirements for memory. Typically, advocates of VI draw attention to fine-grained functional differences, such as the precise shape of Inga's reaction times when recalling information. They argue that these differences involve essential, rather than accidental, properties of memory. Defenders of VE respond that making these fine-grained properties essential to cognitive status commits us to a form of chauvinism about mental life that functionalism was designed to avoid.⁵

In this chapter, we are not concerned to pronounce judgment on the merits of VI versus VE or CI versus CE considered in isolation. Rather, we are interested in whether, on the one hand, taking sides in the dispute over CI and CE implies a commitment to VI or VE, respectively, and on the other hand, whether taking sides in the dispute over VI and VE implies a commitment to CI or CE, respectively. Our primary target is the principle:

INDEPENDENCE

CE and VE are distinct claims that can be accepted or rejected independently.

INDEPENDENCE has generally been assumed to be true by all parties in the disputes above. For example, Rupert, a prominent critic of VE, writes, 'Content external-

4. For the functionalist argument for VE, see Clark (2008); Sprevak (2009); Wheeler (2010).

5. See responses along this line, see Adams and Aizawa (2007); Rupert (2004); Sprevak (2009).

ism and [VE] are distinct, though mutually consistent, theses: neither [VE] nor its negation follows from content externalism, and [VE] does not entail content externalism... I treat [VE] independently of the sort of issues normally addressed in discussions of content externalism' (Rupert 2004, p. 397). Clark, one of the main proponents of VE, agrees: 'In [our original] paper, we showed ... why [VE] was orthogonal to the more familiar Putnam-Burge style externalism' (Clark 2008, p. 78).⁶ In this chapter, we challenge this received view by arguing that INDEPENDENCE is not straightforwardly true. The relationship between, on the one hand, CI and CE and, on the other, VI and VE is more complex than previously suspected. We gestured to VE above, but it turns out to be far from clear how to state VE precisely. Depending on how VE is cashed out, INDEPENDENCE may be either true or false. In the following, we explore some of the intriguing dependencies between VE and CE.

Bear in mind that any branch of externalism is defined as the negation of internalism. It thus follows that if CE and VE are independent, then CI and VI are also independent, and vice versa. Correspondingly, if CI and VI are not independent, then neither are CE and VE, and vice versa. Note, for the record, that we henceforth consider VI/VE and CI/CE claims that could be variously made about one's cognitive, conscious and/or mental life. We also consider them claims that could be variously made about states or processes. A final point to note is that we do not take a stand on precisely where the boundary between the internal and external lies for either VI/VE or CI/CE. We are neutral, for example, about whether 'internal' includes the body, all the nervous system, only the central nervous system or only the brain. Our question is, given a choice for drawing the boundary, what is the relationship between VI/VE and CI/CE?

1 Content externalism

Content internalism (CI) and content externalism (CE) make incompatible claims about the individuation of those mental properties or states which carry content. Individuation is about what makes something what it is. The basic idea of CI is that such contentful properties or states are individuated narrowly, whereas CE takes them to be individuated widely. Consider being a footprint. This is a wide property, because a certain indentation in the sand is a footprint only if caused by a foot. In contrast, the property of being a foot-shaped imprint is narrow in that any intrinsic duplicate of a footprint is a foot-shaped imprint, even if not caused by a foot. Note that both are properties of the sand (or configurations of grains of sand). In particular, the fact that some properties are individuated in virtue of their causal origin does not mean they are properties of those causes. In the case of

6. See also Chalmers (2002).

mental properties or states with wide content, individuation is about patterns of causal relationships. According to CE, an individual can be in a mental state with content only if she sustains appropriate causal relations with her external physical or social environment. The claim is not that every single occurrence of a wide-content mental state has to be caused by certain environmental features. That would be to confuse causation with individuation. Some philosophers use a slightly different terminology to draw essentially the same distinction. Thus, Burge (2010) defines what he calls 'anti-individualism' as the view that 'the natures of many mental states constitutively depend on relations between a subject matter beyond the individual and the individual that has the mental states' (p. 61). That should be distinguished from the claim that the occurrence of a particular mental state causally depends on a subject matter beyond the individual. CI denies that individuals need sustain particular causal relations with their social or physical environment in order to be in a mental state with content. According to CI, such states are individuated solely in terms of features that do not extend into the external environment.

Claims about narrow and wide individuation of contentful mental states are typically cashed out in terms of supervenience. Thus, CI is the claim that the contents of mental states are narrow in the sense of supervening on internal features of individuals who are in those states. Assuming that such states are individuated by their contents, CI amounts to the claim that those states themselves supervene on internal features of individuals. So any two internally identical individuals will also be in the same content-bearing mental states. In contrast, CE is the view that content-bearing mental states are wide in that they fail to supervene on internal features of individuals. Instead, they supervene on the conjunction of such internal features and external features of the individual's social or physical environment. So any two internally identical individuals need not be in the same content-bearing mental states if those environmental features are relevantly different.

So far, content internalism (CI) has been understood to be making a claim about the individuation of content-bearing mental states in terms of internal physical features of individuals who are in those states or, alternatively, about the constitutive dependency of such states on such internal features. Content externalism (CE) would then be the negation of those claims. However, the two notions of individuation and constitutive dependency are not exhaustive. For instance, as we see later, CI and CE can also be characterized in terms of distinct notions of (wide) realization. However, what they all have in common is a commitment to specific supervenience claims. We can thus understand such claims as respective minimal definitions of CI and CE. Thus, consider the following way of cashing out CI in terms of supervenience relations:

SUPERVENIENCE

Content-bearing mental states supervene on internal physical features of individuals who are in those states.

But how exactly should the key notions of ‘supervenience’ and ‘internal’ be understood? Stalnaker (1989) and Jackson and Pettit (1993) emphasize that narrow content should be understood as content shared between internal physical duplicates who occupy the same world: an intrinsic physical duplicate of me need not share my mental contents if located in a world with deviant laws of nature or linguistic practices. To use the analogy above, an intrinsic physical duplicate of a foot-shaped imprint in our world is not itself a foot-shaped imprint if located in a possible world where feet have abnormal shapes. If there is a viable notion of narrow content, it is better to be intraworld narrow than interworld narrow. So the pertinent notion of supervenience should be weak (individual) supervenience, roughly:

SUPERVENIENCE*

States with mental content weakly supervene on internal, physical features P iff necessarily; if individual I_1 is in a state S with content M, then there are some P such that I_1 has P and every other individual I_2 that has P is in state S with M.

Here, internal physical duplicates I_1 and I_2 are content duplicates only if located in the same possible world. Consequently, CE should be understood in terms of failure of such a weak (individual) supervenience claim. Put more positively, CE has it that the supervenience base for states with mental content includes external physical features.

2 Vehicle externalism

Vehicle externalism (VE) appears, on its face, to make a clear and surprising claim about the nature of the mind: the mechanisms of human cognition extend outside the brain and head. On reflection, however, it is not clear exactly what is meant by this claim or whether it really involves a departure from traditional thinking about the mind.⁷ We need a precise formulation of VE. Below, we review four ways of formulating VE. As will be seen, these are not equivalent; they interact in different ways with CI/CE, and they result in different truth values for INDEPENDENCE. The four versions of VE we consider are not meant to be exhaustive, but they do represent some of the principal ways in which VE has been understood to date.

7. For a worry along these lines, see Fodor (2009) and Ladyman and Ross (2010).

The first proposal for stating VE takes its cue from the original description by Clark and Chalmers (1998) of VE as ‘active externalism’:

ACTIVE

VE is true iff an external resource is *active*: the resource is coupled to the agent by a two-way causal loop such that it plays an action-guiding role for the agent in the here and now.

ACTIVE cashes out VE in terms of the presence of a two-way causal loop between the agent and the external resource and characteristic behavioral consequences for the agent with changes to that external resource. Let us look at these conditions more closely.

The first condition requires that the agent’s internal states be not mere causal subjects of the external resource; the agent should be able to modify the resource by causal means. This provides a first contrast between VE and CE. CE, unlike VE, permits external resources to be mere causes for agents. An external resource relevant to CE may lie beyond the agent’s power to control; for example, the distal and historical water samples described by Putnam (1975) affect the content of an agent’s ‘water’ thoughts even though the agent cannot change or causally affect those samples.

Second, VE requires that the external resource guide the agent’s action in the here and now. The relevant sense of ‘action’ is non-intentional; ‘action’ means something like *bodily movement*. This provides a second contrast between VE and CE. A characteristic of CE is that changes to an external resource—for instance, swapping H₂O for XYZ—may change an agent’s intentional content but do not change an agent’s (non-intentionally described) action; the agent would undergo exactly the same bodily movements in both situations.⁸ In contrast, VE requires that changes to, or interventions on, an external resource produce characteristic changes in the agent’s (non-intentionally described) behavior. The effect of these external interventions in a case of VE should be patterned on the effects in a case of neural intervention: ‘If we remove the external component the system’s behavioral competence will drop, just as it would if we removed part of its brain’ (Clark and Chalmers 1998, pp. 8–9).

The second formulation specifies VE in terms of our explanatory commitments:

8. As the so-called slow-switching cases (Burge 1988) illustrate, environmental changes will not immediately result in intentional changes. If you were to be unwittingly transported to Twin Earth, you would begin to think twater thoughts only after you sustain enough causal connections to XYZ (or to other speakers who have interacted with XYZ). Your wide intentional behavior would then change accordingly, e.g. you would reach for twater, where on Earth, when you were thinking water thoughts, you would have reached for water. Still, the physical movements of your arm would remain the same.

EXPLANATORY

VE is true iff an external resource is *explanatorily ineliminable*: one is unable to explain the existence or character of one's mental state/process without making reference to that resource.

Noë uses this formulation of VE. Noë's particular concern is human perceptual experience. He claims that the character of our perceptual experience cannot be adequately explained by neural activity alone; one has also to consider how the brain interacts with the world via bodily knowledge of sensorimotor contingencies. Noë claims that the brain, body and world feature in an explanation of perceptual experience:

I argue that we have reason to believe that the substrates of experience—whatever they are, wherever they are—must be *explanatory substrates*; I argue that the substrates of experience are extended because it is only in terms of non-neural features that we can explain how experience has the character that it does. (Noë 2007, p. 459)

Even if one disagrees with Noë's claim about perceptual experience, one may nevertheless find EXPLANATORY appealing as a way of stating VE. EXPLANATORY suggests that the fortunes of VE are tied to the success or failure of various explanations of mental phenomena. If our explanation of the mind turns out to appeal to extraneural elements, VE is true; otherwise, VE is false. EXPLANATORY is different from ACTIVE. Suppose, for the sake of argument, that reference to an external resource is ineliminable from the explanation of the character of an agent's mental life; that is no guarantee that the same resource also plays an 'active' role for the agent concerned. The resource may be a cause rather than an effect for the agent, and intervening on the resource may fail to change the agent's behavior in the here and now.

A third formulation of VE, suggested by Block Block (2005), uses the notion of a minimal supervenience base:

MIN-SUPERVENIENCE

VE is true iff an external resource is part of the *minimal supervenience base* for that mental state/process.

The minimal supervenience base of a mental state/process is the minimal physical activity needed for that mental state/process to occur. Brain activation of some sort is part of the minimal supervenience base of all our mental processes/states. If there were no brain activation, there would be no mental processes or states. The

question concerning VE is what more, if anything, than brain activation is required for human mental states/processes to occur. MIN-SUPERVENIENCE identifies VE with the claim that external resources feature in this minimal supervenience base. Note that the 'minimal' condition is necessary; otherwise, VE would be trivially true. If brain activity alone were sufficient to produce one's mental states, then brain activity plus activity in an external resource would also be sufficient. For VE to be true, the external resource must play a non-redundant role in the relevant supervenience base. MIN-SUPERVENIENCE differs from EXPLANATORY. There is no reason why explanation of the existence or character of a mental state/process should make ineliminable reference to everything in its minimal supervenience base; indeed, such an explanation is likely to be too detailed to be informative. MIN-SUPERVENIENCE also differs from ACTIVE. Even if an external resource lies inside the minimal supervenience base, that does not guarantee that the resource plays a suitably active role for the agent; many of the neural elements of an agent fail to satisfy ACTIVE's conditions despite being in the minimal supervenience base. A fourth way of cashing out VE uses the notion of realization:

REALIZATION

VE is true iff a mental state/process of an agent is *realized* by the conjunction of the agent's neural activity and an external resource.

REALIZATION identifies VE with the claim that human mental states/processes are realized, not by brain activity alone, but by the conjunction of brain activity and activity in some external environmental resource. REALIZATION has obvious affinity with MIN-SUPERVENIENCE, but it makes a stronger claim. REALIZATION requires not only supervenience on the external resource but also that a particular relation obtain between the mental state/process and that resource—namely, realization. Precisely what this amounts to depends on one's theory of realization. REALIZATION appears particularly apt as a way of understanding VE if one favors extended functionalist arguments concerning VE, since those arguments issue directly in conclusions about the realizers of mental states. REALIZATION differs from EXPLANATORY. The explanation of the existence or character of a mental state/process need not appeal to all, or indeed appeal to only, the realizers of that mental state/process. REALIZATION also differs from ACTIVE. A realizer need not play a suitably active role in the agent's mental life: it need not be subject to the agent's causal control, and interventions on the realizer may not change the agent's behavior in the here and now.

Now that CE and a number of versions of VE are in focus, let us assess whether CE and VE are independent.

3 Assessing INDEPENDENCE

Let us consider how INDEPENDENCE fares under the different formulations of VE.

First, ACTIVE. As we saw above, ACTIVE provides two contrasts between CE and VE: an external resource should be under the agent's causal control for VE but not CE, and changes to an external resource should produce characteristic behavioral changes in the agent in the here and now for VE but not CE. Either of these two conditions entails that a case of CE need not be a case of VE. The converse claim—that VE fails to entail CE—also appears to be true under ACTIVE. Even if VE's conditions are met, mental content may still weakly supervene on internal features of the agent. VE does not entail that if one were to keep the agent's internal state fixed and change the external resource in appropriate ways, the agent's mental content would change. Indeed, the only counterfactuals concerning changes in the external resource entailed by VE concern cases in which the internal physical state of the agent is *not* kept fixed, since the agent's internal physical state must change if her bodily behavior changes. Hence, CE does not entail VE. This is sufficient to establish INDEPENDENCE. An ACTIVE reading of VE is almost certainly behind the claim of Clark and Chalmers (1998) that VE and CE are logically distinct forms of externalism.

ACTIVE allows us to assert that CE and VE are logically independent, but there is a problem with ACTIVE as a general strategy for defending INDEPENDENCE. The problem is that ACTIVE is regarded as an inadequate way of formulating VE. As is seen in Section 2, other formulations of VE take the view that ACTIVE places overly demanding conditions on VE. There seems no reason why an external resource cannot be part of a cognitive mechanism even if that resource only plays the role of a cause, or if changes to that resource fail to produce behavioral changes in the here and now for the agent. Many of our neural resources fail to satisfy ACTIVE's stringent conditions despite playing a role in our cognitive mechanisms. The Parity Principle—a key claim in many arguments for VE—forbids external and internal resources be judged by different standards when deciding their cognitive status.⁹ If one accepts the Parity Principle, ACTIVE cannot be a viable formulation of VE. ACTIVE also suffers from the problem of being too weak as a formulation of VE. Many external resources that advocates of VE do *not* intend as instances of VE satisfy ACTIVE's conditions. For example, the current state of an agent's clothes—whether they are dirty, clean, warm, cold, dry, wet and so on—stands in a two-way causal relation with the agent—the agent will change the state of her clothes if they are in undesirable condition—and interventions that affect the agent's clothes will guide the agent's behavior in the here and now: spilling ink on the clothes will produce

9. See Clark and Chalmers (1998) and Sprevak (2009).

an immediate behavioral response. But just because the agent stands in a two-way causal relation to her clothes which guides her behavior in the here and now does not mean that one has a case of extended cognition.

Second, *EXPLANATORY*. According to *EXPLANATORY*, VE is true just in case one is unable to explain the existence or character of a mental state/process without making reference to an external resource. *EXPLANATORY* supports one half of *INDEPENDENCE* by blocking entailment from VE to CE. That reference to an external resource is ineliminable from our explanation does not entail that CE is satisfied, as it does not guarantee that an agent's mental content fails to weakly supervene on her internal features. Suppose that reference to an external resource, E, is ineliminable from the explanation of the character of agent A's mental life. It could be that E plays no role in individuating A's mental content. E may be ineliminable to explain some other, non-content-individuating, aspect of the character of A's mental life, such as the way in which A processes states with content. Two internal duplicates, A and A*, could have the same mental content and differ in other aspects of their mental lives. Therefore, it is possible for VE to be true without CE being true. This would only be threatened if one had independent reason to think that the identity of an agent's mental content depends on E, or on the aspect of the agent's mental life to which VE pertains. This may happen if, for example, one understands VE as pertaining to how the agent processes mental content and combines this with a version of inferential role semantics.

However, *EXPLANATORY* fails to block entailment in the other direction. An agent's mental content is an important part of her mental life. If two agents differ in mental content, the character of their mental lives should be explained in different ways. Suppose that agent A has a belief about water. According to CE, in order for A to have this belief, and in order for that belief to be about water rather than twater, A needs to stand in an appropriate causal relation to external instances of water. An internal duplicate of A who lacks these causal relations would have a mental life with different mental content, or no equivalent mental content at all.¹⁰ In order to explain the particular character of A's mental life, one needs to make reference to resources outside A—namely, to external water samples. But if reference to an external resource is ineliminable to the explanation of the character of an agent's mental state, then VE is true. Under *EXPLANATORY*, CE entails VE, and *INDEPENDENCE* fails.

Third, *MIN-SUPERVENIENCE*. According to *MIN-SUPERVENIENCE*, VE is true just in case the minimal supervenience base of an agent's mental life includes an external resource. Like *EXPLANATORY*, *MIN-SUPERVENIENCE* supports *INDEPENDENCE* in one direction but not the other. *MIN-SUPERVENIENCE* blocks entailment from VE

10. For duplicates with no mental content, see Putnam (1981), Ch. 1.

to CE. If an external resource, E, is part of the minimal supervenience base, that does not entail that E plays a role in individuating the agent's mental content. E may be required in the supervenience base to fix other aspects of the agent's mental life, such as how she processes her mental content. Mental content can weakly supervene on internal features even if other aspects of mental life do not. Therefore, VE does not entail CE. This would only be threatened if one had reason to think that an agent's mental content is individuated by E, or by the aspect of the agent's mental life to which VE pertains. Certain forms of inferential role semantics may offer this. A more serious problem is that MIN-SUPERVENIENCE fails to block entailment from CE to VE. CE is understood as failure of mental content to weakly supervene on internal features of the agent. Mental content is one aspect of the agent's mental life. If CE is true, then an external resource is in the minimal supervenience base of the agent's mental life. Hence, VE is true. Therefore, INDEPENDENCE fails.

Can one somehow disentangle VE and CE in the supervenience base and save INDEPENDENCE? Intuitively speaking, VE and CE seem to make claims about distinct aspects of the supervenience base. CE is a claim about the supervenience base for content; VE is a claim about the supervenience base for the vehicles that represent that content. The problem, as we saw above, is that SUPERVENIENCE* and MIN-SUPERVENIENCE fail to reproduce this distinction. The challenge is to distinguish between those physical features of the supervenience base that constitute the external vehicle and those that individuate content. We call this the *Demarcation Problem*.

In response to the Demarcation Problem, it might be suggested that we can separate the two aspects of the supervenience base if the notion of minimality in MIN-SUPERVENIENCE is understood in terms of being a metaphysically necessary part of a metaphysically sufficient condition. For example, Block (2005) suggests that a minimal supervenience base should be understood as a Mackie-style INUS condition.¹¹ But an INUS rendition of VE represents no progress in solving the Demarcation Problem. The reason is that CE is cashed out in much the same way. So far, we have emphasized that content externalists include external features as part of the supervenience base for states with mental content. But typically they do not exclude from that supervenience base all internal features. So according to content externalists, a state S with mental content M weakly supervenes on a conjunction of internal physical features P_{int} and external physical features P_{ext} . On their view, it follows that the conjunction $P_{int} \& P_{ext}$ is a sufficient (but unnecessary) condition for state S with content M to occur. That is exactly what the second part

11. Mackie (1965) proposed that talk of causes involves INUS conditions: insufficient but necessary parts of a condition which is itself unnecessary but sufficient for the occurrence of the effect. The condition Block has in mind is different in that a minimal supervenience base for a mental state is distinct from whatever caused that state.

of SUPERVENIENCE* says. In fact, since P_{int} & P_{ext} cannot be a causal condition on the obtaining of state S, it is natural to view that conjunction as a metaphysical condition. Further, since such conjunctions have their conjuncts essentially, hence with metaphysical necessity, P_{ext} will then be a metaphysically necessary part of a metaphysically sufficient condition for S.

The upshot is that if the vehicle externalist opts for Block's INUS-style formulation of VE, she faces the problem that CE is also formulated in INUS-style terms. The Demarcation Problem arises again: how to distinguish those features that constitute the external vehicle from those that individuate wide content. The INUS-style proposal gives no resources to draw this distinction. In particular, whatever external physical features form a metaphysically necessary part of a metaphysically sufficient condition for a state with mental content to exist are bound to include features that form a metaphysically necessary part of a metaphysically sufficient condition for the external vehicle that carries the content of that state. For if the features necessary for the agent to have that content are missing, then the agent's processing of vehicles with that content would simply not occur.

We do not pursue this particular line any further. Instead, we argue that, more surprisingly, the Demarcation Problem afflicts someone who endorses both CI and VE. At first blush, the Demarcation Problem does not seem to arise for this particular combination of views. After all, if we draw the internal/external distinction around the skin and skull, it looks as if the skin/skull boundary could do useful work in distinguishing the relevant features of the supervenience base. We said above that CI involves the claim that mental content weakly supervenes on internal physical features. To say that a physical feature is internal to some individual is to say that it is located inside the skin and skull of that individual. In contrast, VE says that an external resource is part of the minimal supervenience base for the mental state/process in question. To say that a resource is external to some individual is to say that it resides outside the boundary of the individual's skin and skull. So it looks like a friend of CI and VE can avail herself of the skin/skull boundary to solve the Demarcation Problem: the physical features that play a role in individuating the content of mental states are internal to the individual, but physical features pertaining to the vehicles of those mental states include features external to the individual.

While the foregoing looks initially promising, a problem arises. According to VI, the internal/external boundary can safely be drawn around the skin and skull, but once VE is accepted, the boundary between the cognitive system and the external environment may be revised to include whatever external resource—notebook, iPhone or what have you—as an integral part of the cognitive system. Importantly, this consequence is explicitly accepted by content internalists. Here are two illustrative

passages from Chalmers and Jackson, who also both endorse at least the possibility of cognitive or mental extension:

It may even be that in certain cases, epistemic [narrow] content can itself be constituted by an organism's proximal environment, in cases where the proximal environment is regarded as part of the cognitive system: if a subject's notebook is taken to be part of a subject's memory, for example (see Clark and Chalmers 1998). Here, epistemic content remains internal to a cognitive system; it is just that the skin is not a God-given boundary of a cognitive system.

(Chalmers 2002, footnote 22)

... the live issue, and the issue on the table here, is whether or not duplicates from the skin, doppelgangers, in our world might differ in belief by virtue of a difference in their environment. In worlds where people think with major assistance from machines that they plug their brains into, doppelgangers will differ in what they believe (the skin will not be the pertinent boundary). (Jackson 2003, p. 57)

The point is that once the internal/external distinction is redrawn to reflect the cognitive or mental extension of the original skin-and-skull-bound individual, the friend of CI and VE can no longer claim that the physical features that play a role in individuating the content of mental states are distinctively internal to the individual who is in those states or that the physical features that pertain to vehicle externalism are distinctively external to the individual who is in those states. The skin and skull no longer constitute the pertinent boundary. CI will instead assert that the content of mental states supervenes on physical features of individuals that are inside the extended cognitive system—the system comprising the biological organism plus whatever augmenting technological devices serve to extend the mind. But thus understood, the physical features that pertain to the vehicles and the physical features that pertain to the content will both count as internal to the individual whose states they are. Consequently, the Demarcation Problem reappears as a concern about how to separate the internal physical features that play a role in individuating the content of mental states from those internal features on which the vehicles supervene. Once the internal/external distinction is no longer of any avail, it looks as if CI and VE are no better placed than CE and VE when it comes

to responding to the Demarcation Problem.¹²

Finally, let us assess INDEPENDENCE under the REALIZATION formulation of VE. According to REALIZATION, VE is true just in case one's mental states/processes are realized by the conjunction of one's neural activity and an environmental resource. As mentioned in Section 2, the distinctive content of REALIZATION hinges on one's notion of realization. If REALIZATION is to do better than MIN-SUPERVENIENCE in establishing INDEPENDENCE, it must impose more constraints than MIN-SUPERVENIENCE. What might those be? One obvious strategy is to employ Shoemaker (1984, 2007)'s distinction between *core* and *total* realizers and its subsequent elaboration by Wilson (2001) into *narrow*, *wide* and *radically wide* realization. Shoemaker draws the initial distinction as follows:

A total realizer of a property will be a property whose instantiation is sufficient for the instantiation of that property. A core realizer will be a property whose instantiation is a salient part of a total instantiation of it. (Shoemaker 2007, p. 21)

Wilson operates with a trichotomy of core realizations, non-core parts of total realizations, and background conditions. A *total* realization of a realized property is constituted by a *core* realization plus the *non-core part of the total realization*. The *core* realizer of a property is the specific part of the physical system most readily identified as playing a causal role in producing or sustaining the realized property—the role which defines that property. The *non-core* realizer is the part of the system which needs to be activated if the core realizer is to play the causal role in question. The *background conditions* pertain to general features beyond the system necessary for its existence and functioning. The total realizer of a property is then a property of the system, containing any given core realization as a proper part, that is metaphysically sufficient for that property. While the realized property is one that an *individual* has, the *system* need not be identical to that individual. So we must distinguish between the bearer of the realized property and the system whose complete states comprise the total realizer.

Wilson proceeds to define a *narrow realization* as a total realization whose non-core part is located within the individual who has the realized property. In contrast, a *wide realization* is a total realization whose non-core part is not located entirely within

12. There are reasons independent of the debate over VE for thinking that the internal/external distinction should not be drawn around the skin/skull. Take the meningitis example of Farkas (2003). You and I both have symptoms typical of meningitis, but whereas mine are caused by meningitis, yours are caused by a different bacterium. So while we are physically distinct from the skin in, we yet inhabit identical physical environments. Without our knowledge, our token sentences containing 'meningitis' express distinct propositions.

the individual who has the realized property. Finally, a *radically wide realization* is a wide realization whose core part is not located entirely within the individual who has the realized property.

How are we to understand the contrast between CE and VE using this framework? On the one hand, CE appears to be best characterized as a case of wide realization. This fits neatly with the definition of CE offered in terms of weak supervenience. A case of wide realization is one in which weak supervenience on internal features of an individual fails. Another way in which weak supervenience could fail of course is by radically wide realization. And we saw above that both CE and VE violate weak supervenience. What we wanted was to distinguish between *different ways* in which weak supervenience could fail. This appears to be precisely what the contrast between wide and radically wide realization provides. Background conditions appear to offer yet another way in which weak supervenience could fail. However, as seen in Section 1, although background conditions violate supervenience on the individual, they do not violate *weak* supervenience. An intrinsic duplicate of an individual who is in some narrow content mental state need not be a narrow content duplicate if she is in a possible world where different background conditions obtain. So we have:

REALIZATION-CE

CE is true iff the property of having a content-bearing mental state/process is widely realized by the individual's neural activity and an external resource.

VE is naturally characterized as a case of radically wide realization. The vehicle is the core part that most saliently plays the causal role in question. So we have:

REALIZATION-VE

VE is true iff the property of being in a content-bearing mental state/process is radically widely realized by the individual's neural activity and an external resource.

The REALIZATION proposal equates the contrast between VE and CE with the contrast between wide realization and radically wide realization.

Although initially promising, this proposal faces a number of challenges. First and foremost is that the distinction between core and non-core realizers is not of a kind that univocally supports INDEPENDENCE. As both Wilson and Shoemaker emphasize, the core/non-core realizer distinction is interest relative. A core realizer is defined as the part of total realizer that plays a salient causal role in producing or sustaining the realized property. But what makes a contribution salient? And

salient to whom? In certain contexts, some physical resources are salient, in other contexts those elements fade into the (non-core realizer) background. Whether a given physical resource is a salient contributor to the total realizer depends on one's explanatory, descriptive, predictive and other interests. Consider a spring-loaded-bar system, which is a total realizer of the kind *mousetrap* because it fulfills the causal role of catching and killing mice. Suppose one's explanatory interests lie in how this physical system kills mice. Certain features of the spring-loaded bar system will stand out as salient and therefore as core realizers of the *mousetrap*: the high-tension spring, the rigid bar, the sensitive trigger. Other elements will be demoted to non-core realizers. Suppose now that one's explanatory interests change and one wishes to explain how the instance of *mousetrap* attracts mice. Other physical features will stand out as core realizers: the accessible open-air platform, the ripeness of the cheese, the high-friction wooden base. In this explanatory context, the spring, bar and trigger will be demoted to non-core realizers. As Wilson emphasizes, the core/non-core realizer distinction is not an objective and mind-independent distinction but a malleable boundary that is reshaped as a function of our interests.

This complicates the contrast between VE and CE. On the face of it, the core realizers of cognition are the vehicles of cognition with which VE is concerned. In some contexts, this seems to be true. For example, if we wish to explain how Otto is able to recall that there is water on Mars but not on the Moon, the most salient causal contributor to Otto's realizing this property is the physical state of Otto's notebook and his related neural and bodily processes. Similarly, the most salient causal contributors to Inga realizing the same cognitive property are her neural states. In this explanatory context, external content-fixing facts—such as distal samples of water—fade into being non-core realizers. But if our explanatory interests change and we wish to explain how Otto or Inga is able to remember facts about water rather than twater, then according to CE, environmental features *do* play a salient causal role in producing and sustaining the relevant cognitive ability. In this context, content-fixing facts are part of the core realizer. Therefore, in this explanatory context, we also have an instance of radically wide realization. Hence, CE entails VE. INDEPENDENCE is not true or false *simpliciter* but indexed to our explanatory, descriptive, predictive and other interests. INDEPENDENCE may flip-flop from true to false as those interests change.

A second problem with REALIZATION is that it fails to block entailment in the other direction, from VE to CE. Paradigmatic cases of VE involve external resources that include representational states as a salient causal player (e.g., inscriptions written in Otto's notebook). If VE is correct, then these external representational states are among the core realizers of some of Otto's mental states/processes. But these external states cannot play this causal role alone; they need help from other instances of properties in the external environment. For example, in order for the

inscriptions in Otto's notebook to play their role in Otto's mental life, other external supports need to be in place: Otto needs pockets to carry the notebook, Otto needs functioning arms and fingers to use the notebook, Otto may need spectacles to read the inscriptions, and Otto may need a pen to correct entries. In any given case of VE, there is a nexus of additional external property instances that need to be in place for the extended core realizer to play its causal role. These external property instances will be among the non-core realizers of Otto's content-bearing mental state. But if any non-core realizers of Otto's content-bearing mental states extend, then CE is true too.

4 Conclusion

Both advocates and critics of VE have assumed that CE and VE are logically independent. We have found this assumption to be problematic. The relationship between the views is more complex than it first appears. We have seen that the primary reason for this entanglement is due to variation in stating VE. We have examined four ways of stating VE and found that none offer straightforward grounds to accept INDEPENDENCE. We wish to propose that a priority for future work on VE is the formulation of an agreed statement of the view that can be used for evaluating its place in the philosophical landscape.

References

- Adams, F., and K. Aizawa. 2007. *The Bounds of Cognition*. Oxford: Blackwell.
- Block, N. 2005. 'Review of Alva Noë's *Action in Perception*'. *Journal of Philosophy* 102:259–272.
- Burge, T. 1988. 'Individualism and self-knowledge'. *Journal of Philosophy* 85:649–663.
- . 2010. *Origins of Objectivity*. Oxford: Oxford University Press.
- Chalmers, D. J. 2002. 'The components of content'. In *Philosophy of Mind: Classical and Contemporary Readings*, edited by D. J. Chalmers, 608–633. Oxford: Oxford University Press.
- Clark, A. 2008. *Supersizing the Mind*. Oxford: Oxford University Press.
- Clark, A., and D. J. Chalmers. 1998. 'The extended mind'. *Analysis* 58:7–19.
- Farkas, K. 2003. 'What is externalism?' *Philosophical Studies* 112:187–208.

- Fodor, J. A. 2009. 'Where is my mind?' *London Review of Books* 31:13–15.
- Hurley, S. 1998. *Consciousness in Action*. Cambridge, MA: Harvard University Press.
- . 2010. 'Varieties of externalism'. In *The Extended Mind*, edited by R. Menary, 101–153. Cambridge, MA: MIT Press.
- Jackson, F. 2003. 'Narrow content and representationalism, or Twin Earth revisited'. *Proceedings and Addresses of the American Philosophical Association* 77:55–70.
- Jackson, F., and P. Pettit. 1993. 'Some content is narrow'. In *Mental Causation*, edited by J. Heil and A. Mele, 259–282. Oxford University Press.
- Kallestrup, J. 2011. *Semantic Externalism*. London: Routledge.
- Ladyman, J., and D. Ross. 2010. 'The alleged coupling-constitution fallacy and the mature sciences'. In *The Extended Mind*, edited by R. Menary, 155–165. Cambridge, MA: MIT Press.
- Mackie, J. L. 1965. 'Causes and conditions'. *American Philosophical Quarterly* 2:245–264.
- Noë, A. 2007. 'Magic realism and the limits of intelligibility: What makes us conscious?' *Philosophical Perspectives* 21:457–474.
- Putnam, H. 1975. 'The Meaning of "Meaning"'. In *Mind, Language and Reality, Philosophical Papers, vol. 2*, 215–271. Cambridge: Cambridge University Press.
- . 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Rupert, R. D. 2004. 'Challenges to the hypothesis of extended cognition'. *Journal of Philosophy* 101:389–428.
- Shoemaker, S. 1984. 'Some varieties of functionalism'. In *Identity, Cause and Mind*, 261–286. Cambridge: Cambridge University Press.
- . 2007. *Physical Realization*. Oxford: Clarendon Press.
- Sprevak, M. 2009. 'Extended cognition and functionalism'. *Journal of Philosophy* 106:503–527.
- Stalnaker, R. 1989. 'On what's in the head'. *Philosophical Perspectives: Philosophy of Mind and Action Theory* 3:287–316.
- Wheeler, M. 2010. 'In defense of extended functionalism'. In *The Extended Mind*, edited by R. Menary, 245–270. Cambridge, MA: MIT Press.
- Wilson, R. A. 2001. 'Two views of realization'. *Philosophical Studies* 104:1–30.