I

EDITED BY MILJANA MILOJEVIĆ

EXTENDED MIND AND EXTENDED COGNITION

PROŠIRENI DUH I PROŠIRENA KOGNICIJA

UDK: 159.95
https://doi.org/10.2298/FID1804473V
Original Scientific Article
Received: 05.08.2018. Accepted: 26.10.2018.

PHILOSOPHY AND SOCIETY
VOL. 29, NO. 4, 471–646

Gottfried Vosgerau

VEHICLES, CONTENTS AND SUPERVENIENCE

ABSTRACT

In this paper, I provide an argument for the assumption that contents supervene on vehicles, which is based on the explanatory role of representations in the cognitive sciences. I then show that the supervenience thesis together with the explanatory role imply that the individuation criteria for contents and vehicles are tightly bound together, such that content internalism (externalism) is in effect equivalent to vehicle internalism (externalism). In the remainder of the paper, I argue that some of the different positions in the debate stem from different research questions, namely the question about the acquisition conditions and the question about the entertaining conditions for mental representation. Finally, I argue that the thesis of externalism is much more interesting if understood as a claim about how mental representation works in our world as opposed to how they work in all metaphysically possible worlds. In particular, I argue that this "nomological" understanding of the thesis is able to explain how and why the experimental methods used in contemporary cognitive sciences are able to provide insight into behavior generation.

KEYWORDS

externalism, semantic externalism, vehicle extrenalism, supervenience, empirical methods

There are two aspects of each (mental) representation, namely the content and the vehicle, which should not be confused (Hurley 1998). Since there is a clear conceptual difference between the two, theses about them clearly state different things. For example, the thesis of content externalism (Putnam 1975, Burge 1979) and the thesis of vehicle externalism (Clark 2008) are about two very different state of affairs; so different indeed that some claim the two theses to be independent of each other (Rupert 2004). The aim of this paper is to show that, although there is a clear conceptual difference, the two concepts are nevertheless strongly bound together; so strong that claims about the one entail certain views on the other. Of course, all this heavily depends on our understanding of what contents, vehicles and representations are. My understanding of these terms is based on the praxis that we find in behavioral sciences and that led to the so-called "cognitive turn"; thus, I will only focus on understandings of these terms that are compatible with behavioral science praxis.

My argument will proceed in three steps. The first step shortly discusses the notion of representation with respect to its explanatory role as we find it in behavioral sciences. I will argue that representations are assumed (or assigned to cognitive

Gottfried Vosgerau: Professor, Institut für Philosophie, Heinrich-Heine-Universität Düsseldorf; vosgerau@hhu.de.

systems) as theoretical entities that explain flexible behavior, which serves as the first premise of my argument (section 1). The second step establishes the second premise, namely that contents must supervene on vehicles if they are to fulfill this explanatory role. Since there are multiple possible ways of understanding the notions of content, vehicle and representation, I will try to clarify my understanding of them by defending my argument against possible objections that might arise on the basis of different understandings not compatible with behavioral science praxis (section 2). In the third step, I show that (given the premises) content externalism logically entails vehicle externalism, and that vehicle externalism entails content externalism (not strictly logically but in effect; section 3). I conclude that every argument for content externalism is in effect an argument for vehicle externalism and the other way round.

In section 4, I introduce the distinction between acquisition conditions and entertaining conditions as two different explananda for a theory of representation: It is one thing to explain how it is possible to acquire a certain mental representation and another thing to explain what it means to entertain a representation. This distinction is scrutinized in terms of supervenience relations. Then I sketch the debate between Alva Noë and Ned Block, showing that Noë argues mainly for acquisition conditions, while Block offers good arguments regarding entertaining conditions. Although both authors contend to say something about the metaphysics of mental representations (i.e. about the metaphysically necessary conditions for mental representation), I conclude by arguing that externalism is better understood as a claim about the nomological conditions that constrain mental representation. In this way, its impact on behavioral sciences can be much better accounted for.

1 Representation and Explanation

Let me start with the notion of representation.[1] Mental representations are introduced in order to explain behavior. This is at least the idea on which cognitive (behavioral) science is built: There are kinds of behavior, namely flexible behavior, which cannot be explained by stimulus-response patterns. Flexible behavior is understood as behavior that, given one and the same type of stimulus, can still differ. (This general description is meant to include cases in which the behaving system does not have a relevant input at all.) Since this implies that there is no simple one-to-one relation between stimulus and behavior, flexible behavior is not explainable by simple stimulus-response patterns. Rather, or so the reasoning goes, we have to assume that some kind of inner state of the behaving system also has an influence on what kind of behavior is selected given a specific stimulus. These inner states are, moreover, assumed to stand for something else and are hence called "mental representations". In the words of J. Haugeland: "[...] if the relevant features are not always present (detectable), then they can, at least in some cases, be represented; that is, something else can stand in for them, with the power to guide behavior in their stead. That which stands in for something else in this way is a *representation*"

1 I am not going to sketch an account of mental representation here, but rather introduce some basic considerations on why mental representations are introduced, i.e. on what the explanatory role of them is. For an account of mental representation see (Vosgerau 2009).

(Haugeland 1991: 62; original italics). In this sense, we have to assume some kinds of mental representation to explain flexible behavior.[2]

A typical example of (a simple) flexible behavior that is explained by mental representation is the so-called "homing behavior" of the desert ant: The ant is able to go straight back to its nest after an unsystematic search for food (cf. Gallistel 1993). If the ant is displaced before it starts its way back, it will not return to the nest but take a route parallel to the "correct" route that leads it to the point where the nest would have been, had it been displaced in the same way as the ant. This already shows that the ant does not have the relevant features available, i.e. it has no access to relevant features of the nest that could "directly" guide its behavior. Thus, the ability of the ant to return to its nest (i.e. the homing behavior) can only be explained if we assume that the ant has some kind of representation of the location of its nest (Vosgerau, Schlicht and Newen 2008). It is a representation of the location of the nest *because* it explains the ant's ability to find its nest (under normal or favorable circumstances; i.e., errors may occur under certain circumstances). If the ant behaved differently, e.g., if it went to trees instead of its nest, we would have to assume a different representation, namely a tree-representation.[3] A very similar argument was brought forward by R. Cummins against teleological theories of mental representation; his example is: if we observed bee dances that indicate the location of piles of rocks, than we would have to assume that they represent the location of piles of rocks, "even though we would be mystified about the evolutionary significance of the whole business" (Cummins 1989, 86). Therefore, the content of representations is, in the first place, determined by the object of the behavior to be explained. (There might be further criteria for content individuation, e.g., Frege's principle of simultaneous believability [Frege 1892]. However, such further criteria can never tell us why it is a nest-representation rather than a tree-representation.) In other words: If the ant would not interact[4] with its nest, we would have no reason to assume that it has a nest-representation at all.

The point to be made here is not just epistemological but rather conceptual: The concept of representation can only be applied to things that stand for something (the represented entity). "Standing for something" means (at least in the case of mental representations) that the representation can be used by a behaving system to engage in behavior directed towards the represented entity without having direct sense-contact with it.[5] This means that the representation enables the behaving system to act as if it had direct sense-contact with the represented entity.

2 This idea can also be found in the pioneering work that led to the assumption of mental representations, e.g. in Tolman and Honzik (1930), Tolman (1948); (cf. Vosgerau 2010).

3 The formulations just given might be read to point to a dispositional understanding of mental representations. Because of space limitations, I cannot discuss this idea—which I reject—here. For a detailed discussion see Vosgerau (2010).

4 I use "interaction" in a wide sense, which includes avoidance behavior. E.g., when a beaver flees from an eagle, it interacts with the eagle in this sense; it displays, as I will say, an eagle-directed behavior, and the eagle is the object of the fleeing-behavior.

5 This is true at least for perceptual representations. There might be further (conceptual) representations that are built by combining representations. Such conceptual representations can be far more abstract in that they might represent things that cannot be the target of behavior. However, it is very plausible to think that such conceptual representations are based on perceptual representations (cf. Vosgerau 2009).

Crucially, the basic idea is still that representations fulfill a certain explanatory role: *Because* representations are substitutes for the represented entities, the bearer of the representation is able to show this and that behavior. In a full account, however, this basic principle has to be spelled out in more detail, for example by explaining the explanatory roles in terms of causal roles. Nevertheless, on a more abstract level of description, the content of the representation, which specifies the represented object, will explain the behavior. In the case of the ant, the content of the representation can be described as location of the nest, and it enters the explanation of the ant's homing behavior at the place where the location of the nest (or the detection of the relevant features of the nest, resp.) would occur if the ant had sense-information about the nest.

2 Supervenience as the Relation Between Contents and Vehicles

Let me now turn to the second step in my argument, namely the defense of the claim that contents supervene on vehicles. The distinction between contents and vehicles is widely accepted, and it is widely agreed that we should not confuse the two (Hurley 1998). Accordingly, a strict difference between content externalism (going back to Putnam 1975, Burge 1979) and vehicle externalism is drawn, since they amount to radically different claims (Rupert 2004). While I agree that these are, in principle, two different issues, the question arises how they are connected. Is it, for example, possible to hold the one but to deny the other? The answer to this question depends upon one's view on content and vehicle individuation: If contents are individuated independently from vehicles and vice versa, it is possible to separate the two issues clearly. But if the individuation of vehicles is dependent on the individuation of contents, theses about the "location" of contents have implications for the location of vehicles. In this paper, I argue that there is a systematic relationship between content individuation and vehicle individuation, namely the supervenience of contents on vehicles.

Contents are introduced in order to explain behavior (see above). Such explanations are usually understood as causal explanations ("the ant goes back to its nest because it has a representation with this and that content"). If we now ask what the vehicles of these contents are, it seems quite natural to say that contents supervene on their vehicles. My argument for this claim takes the form of a *reductio*: Firstly, vehicles are things that play causal roles in the behaving system that displays the behavior. Now assume that the content of those vehicles does not supervene on them; then it would be possible that one and the same vehicle has different contents (this follows from the definition of supervenience). This would mean that there could be two different content-based explanations for a certain behavior (which involves one and the same vehicle) for which there is only one causal story on the vehicle level. From this we can infer that the contents of mental representation would not be apt to figure in causal explanations of behavior, i.e. they would have no causal roles. This, however, contradicts our idea of providing causal explanations of behavior based on content. From this contradiction we can infer that the assumption was wrong, and therefore, that contents supervene on vehicles.

Of course, contents are but one of the causes of behavior among other causes. This could lead to the following reasoning: Contents are based on (or involve) more

causal roles then just the causal role of the vehicle, i.e. content is individuated externalisticaly and vehicles are individuated internalisticaly. Take, e.g., the mental state of seeing a tree. It could be argued that the tree plays a causal role relevant for the individuation of the content of this mental state, yet the vehicle is the according brain state that is internal. How then would the vehicle be individuated? It could be individuated as the internal part of the supervenience base of the content. However, this criterion for vehicle individuation is rather ad hoc and does not seem to have an independent support (indeed, it would dogmatically presuppose vehicle internalism). Moreover, it would give us two different causal explanations: one on the content-level and a different one on the vehicle-level.[6] And what then would make this internal state the vehicle of this content (rather than another content)? There is no principle in sight which could answer this question. So, if we want to construe the content of seeing-a-tree externalistically (i.e. construing "seeing" as a success-verb), we should take its vehicle to involve the tree as well; if we want to talk about the brain state as the vehicle, we should construe the according content as having-the-impression-of-seeing-a-tree (we could be mistaken after all), which is individuated internalistically as well.

It might be argued that my reductio is way to simple, given the sophisticated arguments by Burge (1979) to the contrary effect. But, Burge's arguments for content externalism do not contradict my claim, since Burge merely establishes the thesis that the content of mental states does not supervene on brain states (or, more generally, inner states of the representing system). However, he does not claim that brain states are the vehicles of mental content. Therefore, his claim is compatible with the claim that contents supervene on vehicles. It is only incompatible with the claim that brain states are vehicles: "CE [content externalism] has it that the supervenience base for states with mental content includes external physical features" (Sprevak and Kallestrup 2014: 82).

Furthermore, Burge's argument (the twin earth argument) is a variation of Putnam's (1975) argument for externalism of linguistic meaning. Burge adds the premise that the content of mental states is individuated by (or even identical to) the content of the linguistic expression we use to describe the mental state. This notion of mental content is, however, quite obviously not compatible with behavioral science praxis: The explanation of the homing behavior of the ant is not dependent on how we describe its mental representation. So, if we want to go along with cognitive science and ascribe a mental content to the ant that has a role to play in the explanation of the ant, we better not use a notion of content that is language-dependent such as Burge's notion (cf. Newen and Vosgerau 2007).

Another line of reasoning might be that not only content but also behavior could be individuated broadly or narrowly. And if so, content could play a role in explaining broadly (i.e. externalistically) individuated behavior, while the vehicle only plays a role in explaining the narrowly (i.e. internalistically) individuated behavior. In this way it might be that vehicles are internal and contents external, implying that contents do not supervene on vehicles. Again, the notion of broadly

6 While it is plausible that content explanations might be on a different level of abstractness (or fine-graindness), we still expect a systematic correspondence between the two levels if both are considered to deliver causal explanations (Soom 2011).

individuated behavior is at odds with the praxis of behavioral sciences: It is crucial that the experimentally induced cases, in which the ant does not reach its nest, are equally subsumed under the same type of behavior, namely the homing behavior. If behavior would be individuated broadly, these cases would have to count as cases of a different type of behavior (because it does not involve the nest), in which case they could not give us any evidence about the case we are interested in, namely the ability of the ant to find its nest. Thus, the notion of behavior that is used in the behavioral sciences has to be the narrowly individuated one.

Sprevak and Kallestrup (2014) claim that at least content externalism has to be confined to a claim about weak supervenience instead of strong supervenience (Sprevak and Kallestrup 2014: 82), building on arguments from Stalnaker (1989) and Jackson and Pettit (1993) concerning narrow content. And if so, then it simply is not true that one and the same vehicle cannot have different contents: This is only excluded within the same possible world but not in two different possible worlds. However, the argument given for the individuation of narrow content cannot be transferred to content externalism, at least not to the Burge-style content externalism they, and I, have in mind: The Twin-Earth story involves two different possible worlds; thus, it would not tell us anything about the supervenience base for contents then, especially not that "the supervenience base for states with mental content includes external physical features" (Sprevak and Kallestrup 2014: 82). In particular, it would cease to be an argument for the thesis that mental content does not supervene on brain states. Thus, I will proceed with the assumption that strong supervenience is the better option for content externalists.

Last not least, let me shortly comment on the claim that vehicle externalism is the claim that "the mechanisms of human cognition extend outside the brain and head" (Sprevak and Kallestrup 2014: 83). Interestingly, later in the paper the relevant formulations of the different versions of vehicle externalism include reference to, e.g., "mental processes/states" (Sprevak and Kallestrup 2014: 85). Although I do not wish to argue that this understanding is somehow wrong or inadequate, I would like to point to a few problems that lead me to adopt a more specific notion of vehicle than the one implied by the passages above. In particular, this will be the notion that is referred to earlier in their paper, namely the following: "The vehicle of content is the physical item that has, or expresses, that content – for example, a sentence, if we talk about linguistic content" (Sprevak and Kallestrup 2014, 78). Clearly, this is a state, not a process. The process of writing does not have or carry or express the content, rather the ink pattern on the paper does. And in parallel, it is some physical states (e.g. brain states) that are the vehicles of mental content, and not "some piece of cognitive architecture" (Sprevak and Kallestrup 2014: 78), let alone mental processes.[7] And while there is a version of vehicle externalism that talks about processes (playing Tetris and mental rotation, e.g.), my concern here is only with mental states, i.e. with the version of vehicle externalism that is about the notebook entry (not the writing down) of Otto, for example. And this is because my aim is to discuss the notion of mental representation that is used in the

7 Indeed, it seems pretty unclear what would make a mental process mental anyhow (cf. Fodor (2009) and Vosgerau (2010)), which is probably also reflected in the debate about a "mark of the mental" (Walter 2010).

behavioral sciences (see above), which is about the contentful states of animals. So, there are mental states that have two aspects: a content and a vehicle that has or carries the content. And for this understanding of mental states (representations), which I claim to be the relevant one for the praxis of behavioral sciences, I contend that my argument for supervenience holds.

3 The Relation between Vehicle and Content Externalism

I therefore take it that contents supervene on vehicles (if they figure in causal explanations). According to the standard formulation of supervenience (Kim 1984), properties of kind A supervene on properties of kind B iff: $\Box \forall x \, \forall F \in A \, [Fx \rightarrow \exists G \in B \, (Gx \wedge \Box \, \forall y (Gy \rightarrow Fy))]$[8]. In words: Necessarily, if something has A-property F, then there is a B-property G that it also has and, necessarily, anything else having G also has F. For our case, content types (the property of being an instance of this content type) are A-properties while vehicle types (the property of being a vehicle of this type) are B-properties. So, if contents supervene on vehicles, it is impossible that one and the same vehicle has two different contents. At the same time, it is possible that one and the same content has different vehicles ("multiple realization").

Equipped with the two premises (that representations explain flexible behavior and that contents supervene on vehicles), we are now able to move to the third step of the argument, which investigates the relation between content externalism (internalism) and vehicle externalism (internalism). To show that this relation is in effect an equivalence relation, I will discuss both directions of the biconditional in turn.

Assume that content externalism is true, i.e. that at least some contents essentially depend on factors in the environment of the behaving system. Then it is possible to change a content by changing some factor in the world while keeping the state of the behaving system fixed (exactly this is done in the Twin-Earth thought-experiments). Then, there are two different contents, *F1* and *F2*, for each of which there has to be some vehicle on which it supervenes. Since these two vehicles have to be distinct according to the supervenience definition (otherwise, the second implication would be false), and since the internal states of the behaving system are not distinct, the vehicles have to extend beyond the boundaries of the system into the world. Thus, if contents supervene on vehicles, content externalism implies vehicle externalism.[9] (By *modus tollens*, vehicle internalism implies content internalism.)

What about the other direction: does vehicle externalism imply content externalism? Hurley (1998: 3) argues that multiple realizability is the reason why not:

8 This is the standard definition for strong supervenience which is sufficiently adequate for the present discussion. Other versions of supervenience either are not apt for the present discussion or are designed to minimize certain difficulties which are not important for the present discussion. (See also above and the following footnote.)

9 There are attempts to formalize the notion of supervenience such that extrinsic properties like externalistically individuated content can be handled (e.g. Hoffmann and Newen 2007). However, the main point in these formalizations is to find a general constraint that excludes properties from the supervenience base that are irrelevant for the higher-order property; so, these accounts can be said to provide a formalization of the notion of a "minimal supervenience base". The presented argument is, however, independent of how exactly the minimal supervenience base is determined.

While there might be cases in which a content supervenes on internal states, this does not mean that all instances of this content do. In the definition, this fact is expressed in the existential quantifier: We do not know how many *G*s there are, and, for a given content *F*, some *G*s may involve external features while others do not. The reason for this possibility is, according to Hurley, that vehicles are token-explanatory while contents are type-explanatory. Thus, for different tokens of the same (content-) type, there can be different features playing an explanatory role.

However, this picture does not work at least for the purpose of explaining behavior. The reason is that the behavior which we want to explain is always a type of behavior. It might be (metaphysically) possible that the vehicle of the ant's nest-representation involves the nest sometimes (maybe even exactly the times at which the homing behavior is successful), and does not involve the nest at other times (namely in cases of misrepresentation). But how do we find out about this? We cannot, since the explanations we provide are type-explanations. This is just the way experiments work: Different factors are systematically controlled while the single tokens of behavior are counted as belonging to one type—the type that we call "the behavior" (e.g. the homing behavior of the ant). If we would not count cases of (experimentally induced) misrepresentation as belonging to the same type of behavior, then the experimental conditions could not tell us anything about the behavior we want to explain. And so in explaining "the" homing behavior, we have to assume that the nest does not play an essential part because it is not involved in all situations in which the behavior is displayed. If we are really interested in explaining one successful token of the homing behavior, we still resort to the general pattern which we explored for the type. Thus, there is no way of formulating a criterion for individuating the vehicle for this token without reference to the type: There is no reason to assume that in the successful case the internal features of the ant are radically different from the (experimentally induced) case of misrepresentation. Therefore, the explanation we give for tokens relies on the type-explanation we have given. In other words: If, like Hurley proposes, vehicles are token-explanatory, then we need individuation criteria independent from content (since content is type-explanatory); however, there are no such criteria,[10] and if there were, they would not help to explain behavior.

This point can be demonstrated more formally with the help of the notion "minimal supervenience base". The definition of strong supervenience is formulated such that it is always possible to include in *G* every arbitrary feature for any token of *F* (the "irrelevant feature problem"; cf. Hoffmann and Newen 2007). However, what we are interested in is the minimal supervenience base, i.e. the smallest set of *G*s (which does not contain irrelevant features) which renders the definition true (for some specific *F*). Cases of misrepresentation, in which the represented object is not present (or does not have the represented features), show that the minimal supervenience base for these cases does not include features external to the behaving system. This provides very good reason to think that the minimal supervenience base in cases of successful representation is internal, too. Indeed, this is the rationale of experimental research (as discussed above). If there is no case of

10 Vehicles could be a natural kind with essential properties for which the reference is fixed by ostention. This, however, seems to be a very mysterious claim.

misrepresentation, then the "minimal supervenience base" is likely to include external features—however, these cases are not cases of mental representation since the displayed behavior is not flexible. For this reason, cases of misrepresentation constitute very good cases for internalism (pace Hurley).

One might argue that further external features are involved in the ant's representation since it has been shown that the ant represents the direction of the way to the nest relative to the location of the sun. However, this only leads to a more specific characterization of the content of the ant's representation: location of the nest in terms of the direction relative to the sun (and distance). Although this implies that the ant has to have information about the location of the sun in order to successfully use the representation, it does not follow that the sun is part of the minimal supervenience base of this content. However, such a "relativization" of the content is not possible for the nest itself, and thus the nest is not only excluded from the minimal supervenience base, but it also does not play a role in the explanation of the behavior (unlike the location of the sun).

To conclude, vehicle externalism does not imply content externalism logically. However, conjecturing that for some tokens the vehicles are (partly) external (although the content is fully internal) leads to a non-testable hypothesis that contradicts experimental praxis and scientific explanation. Such a thesis would claim that the vehicles of misrepresentations are systematically different from the vehicles of successful representations with the same content. The reason for its non-testability is that tokens of vehicles cannot be individuated independently of content, so that the same criteria for vehicle individuation apply to both successful and misrepresentations. I therefore conclude that every sensible version of vehicle externalism goes hand in hand with content externalism and vice versa. Therefore, the sharp distinction between content and vehicle externalism is very helpful at the conceptual level, but it is not very important in arguing for one or the other.

4 Acquisition and Entertaining of Representations

Indeed, philosophers arguing for externalism (for some mental representation, e.g. perceptions) usually claim that the external features play an *essential* role for representation, i.e. that in every case of representation, the external feature is involved. Combining this with the thesis that content supervenes on vehicles, this kind of argument is both an argument for vehicle as well as for content externalism. As an example I will discuss Noë (2005), who argues that the vehicles of percepts extend beyond the boundaries of the perceiving system. He claims that the perceived object plays an essential part in constituting the percept, because percepts are based on "knowledge" about the systematic contingencies between one's own movements and the changes in the sense input coming from the perceived object.[11]

11 Noë (2005) sometimes seems to make the more moderate claim that only parts of the body are an essential part of representations because movements are (and thus representations are) in the body as opposed to only in the brain. However, the systematic contingencies do in fact involve the environment, such that I will concentrate on this claim. Moreover, he suggests that every occurrence of a percept has to involve an active component—a claim which I will not discuss here (for a critical discussion see, e.g. Block 2005, Jacob 2006).

Block (2005) criticizes this position by claiming that Noë does not talk about the minimal supervenience base. He holds, contrary to Noë, that in the end, vehicles and contents have to be internalistically individuated, even if the represented objects play an essential part in the *normal* way of acquiring the representations in question.

In order to evaluate both positions and the involved arguments with respect to the explanatory role they assign to mental representations, I will first introduce the distinction between the acquisition conditions and the entertaining conditions for mental representations: It is one thing to explore the conditions that have to be fulfilled to acquire a certain representation, and another thing to explore what conditions have to be fulfilled to entertain a given content (provided that the system has acquired this content). With this distinction at hand, I will show that the debate between Noë and Block arises only because they are talking about different things; there is no real controversy. Before doing so, I will now introduce this distinction in detail.

4.1 Historical Externalism

Consider, again, the homing behavior of the ant: The ant has a nest-representation (rather than a tree-representation) because it displays nest-directed behavior. If there were no nests, there would be no nest-directed behavior, and thus there would be no nest-representations. In other words, the ant could never perform any nest-directed behavior and *a fortiori* could not acquire the nest-representation, if it had never interacted with nests.

Since the ant can only have a nest-representation if it has had some interaction with nests, the supervenience-base of the nest-representation has to involve external features: It is always possible that there is an ant with the same internal state as the ant that is displaying homing behavior, but which is living in a world without nests and thus does not have a nest-representation. Therefore, the internal state of the ant cannot function as the *G* in the definition of supervenience, since there are some worlds in which the last implication is false. So, the supervenience base *G* has to contain past interactions with nests. Hence, the supervenience base has to include nests (at least past interactions with nests) and so extends beyond the boundaries of the ant.

Independent of the specific story about perceptual states told by Noë (2005) (but fully compatible with it), mental representations are externallistically individuated (both the contents and the vehicles) in the following sense: It is not possible to display a certain behavior and, *a fortiori*, to acquire a certain mental representation without having had interactions with the represented object. The point is not that the represented object has to be there every time (there are cases of misrepresentation), but that there must be a certain history of the representing system in which it successfully acquired the representation in question. I will call this version of externalism "historical externalism" (authors like Dretske, Lycan, and Tye have defended such an externalism for contents), which follows from the supervenience thesis of content and vehicles if we take into account the acquisition conditions for content, which corresponds to evaluating all possible worlds to interpret the necessity operators in the supervenience definition.

4.2 Entertaining Content

If we aim at fully accounting for the acquisition conditions for mental representations, then it seems that we are forced to adopt historical externalism. However, if we abstract from the acquisition conditions, assuming that the behaving system in question already possesses the ability to entertain a this-and-that-representation, we can focus on the entertaining conditions: What is required of a behaving system to entertain a certain representation (given that it has acquired the ability to do so)? For this end, let us restrict the set of possible worlds which we take into account in evaluating the definition of supervenience. The basic idea is that we consider only worlds in which the system in question has had interactions with the object in question. Since we are dealing with temporal relations, I will use standard temporal modal logic in which a possible world is a world at a certain time-point (whatever a time-point may be).

Let us introduce the notion of an α-world for a behaving system *s*, which is a world in which *s* possesses the ability to entertain an *A*-representation. Such a world has to be preceded by at least one world in which *s* has had interaction with *A*s (within a certain time range).[12] Thus, for every α-world of *s*, the following must be true: $\exists t_i \left[t_{\alpha-\epsilon} < t_i < t_\alpha \wedge V\left(t_i, \exists x\left(Ax \wedge I(s,x)\right)\right) = True\right]$.[13] At least in one world i_i within a certain temporal range ϵ before t_α, the sentence "some *A* exists with which *s* interacts" has to be true. We can now define the set of α-worlds for *s*:

$$\mathcal{W}_\alpha^s = \{t_j \mid \exists t_i\, [t_{j-\epsilon} < t_i < t_j \wedge \mathcal{V}\left(t_i, \exists x\left(Ax \wedge I(s,x)\right)\right) = True]\}$$

Abstracting from the acquisition conditions now means to consider only α-worlds for evaluating the definition of supervenience. (Of course, due to the use of temporal logic to define the sets of possible worlds, we are not allowed to use implicit temporal predicates like "…has had interactions with …" in the object language.) If we do so, the argument for historical supervenience does not work anymore, since now there is no possible world left in which the system fails to have the right acquisition history. And, as we know from cases of misrepresentation, the actual entertaining of a representation does not imply the existence of the represented object. So, if we only look at possible worlds (i.e. worlds at a certain time-point) in which *s* has acquired an *A*-representation (by reducing the set of relevant worlds to those with the right kind of acquisition history), then the presence or absence of *A*s does not change the content of the representation anymore. Thus, there are possible α-worlds in which there is no *A* although *s* entertains an *A*-representation. Accordingly, *A*s cannot be included in the minimal supervenience base if we abstract from the acquisition conditions. Thus, if we focus on the conditions of entertaining a representation, we will have to take cases of misrepresentations seriously and we will thus conclude that both the vehicle and the content are individuated internally. This point seems to be made by Block (2005) when he distinguishes between

12 The limitation to at least one world is only for simplicity of the formula; in fact, it is plausible to assume that the system has to interact with the object several times.

13 I employ the standard notation of temporal logic, where t_i is the world at time-point t_i, and $\mathcal{V}v$ is the evaluation function that maps worlds and formula onto truth values. Moreover, the two-place predicate $I(\nu,\mu)$ is used to express that ν interacts with μ.

causal influence and constitutive factors. He claims that everything that is constitutive for *entertaining* a certain mental representation is internal. At the same time, he does not contradict the idea of historical externalism, presumably because it is not concerned with the constitutive factors for *entertaining* a representation but with the constitutive factors for *acquiring* a representation.

4.3 Metaphysical or Nomological?

The debate about externalism is no doubt a debate about the metaphysics of mental representations. So, advocates of externalism take their claim to be a metaphysical claim (understood as a claim concerning the nature of mental representations) and are thus confined to a metaphysical reading of the necessity operator. In this section, some core arguments will be discussed regarding this claim.

Noë (2005) cites the cat experiments from Held and Hein (1963) to support his externalistic claim. In this experiment, two kittens were put in a vertically striped cylinder. One was allowed to actively explore its surroundings, while the other was passively moved around in the same way the active kitten moved (such that the visual input for both was identical). They could show that the passive kitten had severe deficits in perception compared to the active kitten. Thus, it seems that active exploration of (interaction with) the environment is necessary for the acquisition of perceptual representations. Block (2005) argues that this experiment does not show that active exploration is metaphysically necessary: His counter-argument is that if in the passive kitten the same brain processes would be going on (however they came about), it would be absurd not to believe that the passive kitten had perceptions.

Let us assume that the argument is a good argument against the metaphysical necessity of a specific kind of acquisition history. However, does this mean that it is also a good argument against nomological necessity, i.e. the claim that given the "natural laws" of our world, representations can only be acquired in this way? How could it be that the passive kitten has the same brain states, if not by chance? We could imagine that the passive kitten's brain is connected by wires with the active kitten's brain in an appropriate way. However, in this case it (presumably) would have to be so wired that the motor commands of the passive kitten are also able to control the active kitten. Then, at least two problems arise: (1) Which kitten does effectively control the movements of the active kitten? (2) If the passive kitten could be said to control the movement of the active kitten and to have the perceptual input of the active kitten, then all requirements for having perceptual representations are met, according to the "enactive approach"—we would just be confronted with the weird situation that the brain which processes the relevant signals is at a physical place different from that of the acting body. Thus, Block's argument (Block 2005) cannot show that active exploration of the environment is not nomologically necessary for acquiring representations. (It probably was not intended to show that anyway.)

If the passive kitten had the same brain states by chance, then this thought experiment is essentially the same as the famous "swampman" experiment by Davidson (1987), in which by coincidence a physical twin of Davidson comes into being. The question is whether this swampman, which of course behaves exactly like Davidson, has mental representations or not. According to the picture of mental

representations drawn above, we would have to admit that swampman indeed has mental representations, since he shows flexible behavior which we could not explain otherwise. So, if swampman is a serious metaphysical possibility (and there is not much reason to doubt that, although there is much reason to doubt that it is nomologically possible), then the specific way of acquiring a certain representation cannot be a metaphysical necessity for entertaining this representation. The necessity operator in the supervenience definition therefore has to be understood in terms of nomological necessity, at least for the case of historical externalism. (And in the case of entertaining conditions, the nomologically necessary acquisition condition is already built into the definition of the possible worlds under consideration.)

Thus, as far as the metaphysical conditions for being a representation are concerned, I think that there are always defeating arguments against externalism. It is always metaphysically possible that a physical duplicate of some behaving system is created by chance. There is, by presupposition, no reason to think that this newly created system would behave differently from the "original" system. And, since mental representations are to explain flexible behavior, we have to attribute to these newly created systems the same representations which we ascribe to the original system. Therefore, metaphysically spoken, externalism is wrong. However, it can still be a nomological necessity. The conclusion that we draw from this result depends on the question we ask. If we are interested in the metaphysics of mental representations, the final result is internalism. However, I think that there are good reasons to be even more interested in the results of the nomological discussion: If we want to understand how empirical research works and how it can possibly shed light on mental representation, we should care much more about nomological necessity than about metaphysical necessity. Insofar historical externalism is nomologically true, evolution and psychological development play a major role in the explanation of the acquisition of mental representations. And since the entertaining conditions for mental representations are internal, we can study mental representations in behaving systems by neuroscientific methods that do not account for external factors. The most important conclusion of this paper is hence that the most fruitful debate for the understanding of mental representation is not the metaphysical debate about externalism and internalism, but the debate about the nomological conditions of mental representation, which has direct impact on our way of investigating mental representations in the empirical sciences.

5 Conclusion

Representations are assumed in order to explain flexible behavior. The idea that contents supervene on vehicles is defended on the grounds that otherwise, the explanatory role of representations and/or the praxis of experimental behavioral sciences could not be secured. It then turns out that content externalism implies vehicle externalism. Although vehicle externalism does not logically imply content externalism, it is nevertheless implausible to defend vehicle externalism and content internalism at the same time, since this position (1) has no individuation criteria for vehicles to offer and (2) is not able to do justice to the explanatory role of vehicles in the explanation of behavior. I conclude that arguments for or against externalism are always arguments for or against both vehicle and content externalism.

Taking the debate between Noë (2005) and Block (2005) as an example, I have shown that possibly the two are talking about different things. My proposal is to distinguish acquisition conditions for mental representations from conditions of entertaining a certain representation. While the arguments of Noë (2005) are (plausibly) concerned with acquisition conditions, the counter-arguments of Block (2005) (notedly) discuss entertaining conditions. While it is very plausible that the acquisition of representations does indeed depend on a certain learning history that essentially involves (parts of) the environment, the entertainment of a certain representation does not rely on the environment (provided that the system in question has acquired the ability to entertain such a representation). If we look at all possible worlds, there might always be a behaving system having the same internal state as one that has a certain representation *r*, but that fails to have *r* as well, because it has never interacted with instances of the represented object and so cannot even exhibit the same kind of behavior. Therefore, we have to include external factors in the minimal supervenience base for contents (historical externalism). However, if we restrict the set of possible worlds that we evaluate to those worlds in which the system in question has the right acquisition history, then there is no reason to include external factors into the minimal supervenience base. The reason is that there are always cases of misrepresentation in which the represented object (or represented aspects of objects) are not present. Hence, when we focus on the acquisition conditions for mental representations, we should adopt a form of historical externalism; however, if we focus on the entertaining conditions for a certain representation, then we have good reasons to defend an internalistic view. Moreover, historical externalism is plausible only if we interpret the necessity operator in the supervenience definition as indicating nomological necessity, which might be even more interesting with respect to empirical work.

I have tried to approach some of the debates about externalism from a point of view that takes seriously the experimental praxis of behavioral (cognitive) science. In particular, I have argued that some notions of content and vehicle that we find in the literature on externalism are not apt to do justice to this praxis and to scientific explanation of animals' behavior. If we concentrate, however, on this praxis, a lot of possible confusions about contents and vehicles can be avoided. Moreover, the discussion of the kind of necessity that is involved in supervenience claims showed that understanding externalism as a claim about how mental representations work in our world (as opposed to a claim about the metaphysics of representation being true in every possible world) is much better suited to connect to empirical research. In particular, it is then possible to explain how and why the experimental methods used today are able to provide insight into behavior generation in cognitive systems.

References:

Block, N. (2005), "Review of Alva Noë, Action in Perception", *The Journal of Philosophy* 102: 259–272.

Burge, T. (1979), "Individualism and the Mental", *Midwest Studies in Philosophy* 4: 73–122.

Clark, A. (2008), *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford University Press: Oxford.

Cummins, R. (1989), *Meaning and Mental Representation*, The MIT Press, Cambridge MA: London.

Davidson, D. (1987), "Knowing One's Own Mind", *Proceedings and Addresses of the American Philosophical Association* 60: 441–458.
Fodor, J. (2009), "Where is my Mind? Review of Supersizing the Mind: Embodiment, Action and Cognitive Extension by Clark, A.", *London Review of Books (Online)* 31: 13–15; http://www.lrb.co.uk/v31/n03/jerry–fodor/where–is–my–mind, last access: 2016–03–31.
Frege, G. (1892), "Über Sinn und Bedeutung", *Zeitschrift für Philosophie und philosophische Kritik* 100: 20–50.
Gallistel, C. (1993), *The Organization of Learning*, The MIT Press, Cambridge MA.
Haugeland, J. (1991), "Representational Genera". *In:* W. Ramsey, S. Stitch and D. Rumelhart (eds.), *Philosophy and Connectionist Theory*, Hillsdale: Erlbaum, pp. 61–89.
Held, R. and Hein, A. (1963), "Movement-Produced Stimulation in the Development of Visually Guided Behavior", *Journal of Comparative and Physiological Psychology* 56: 872–876.
Hoffmann, V. and Newen, A. (2007), "Supervenience of Extrinsic Properties', *Erkenntnis* 67: 305–319.
Hurley, S. (1998), "Vehicles, Contents, Conceptual Structure, and Externalism", *Analysis* 58: 1–6.
Jackson, F. and Pettit, P. (1993), "Some content is narrow".In: J. Heil and A. Mele (eds.), *Mental Causation*, Clarendon Press: Oxford, pp. 259–282.
Jacob, P. (2006), "Why Visual Experience Is Likely to Resist Being Enacted", *Psyche* 12: 1–12.
Kim, J. (1984), "Concepts of Supervenience", *Philosophy and Phenomenological Research* 45: 153–176.
Newen, A. and Vosgerau, G. (2007), 'A Representational Account of Self-Knowledge', *Erkenntnis* 67: 337–353.
Noë, A. (2005), *Action in Perception*, Cambridge, MA: MIT Press, London.
Putnam, H. (1975), "The Nature of Mental States". In: *Mind, Language, and Reality*, Cambridge: Cambridge University Press, pp. 429–440.
Rupert, R. (2004), "Challenges to the Hypothesis of Extended Cognition", *The Journal of Philosophy* 101: 389–428.
Soom, P. (2011), *From Psychology to Neuroscience. A New Reductive Account*, Frankfurt: Ontos Verlag.
Sprevak, M. and Kallestrup, J. (2014), "Entangled Externalisms". In: M. Sprevak and J. Kallestrup (eds.), *New Waves in Philosophy of Mind*, New York: Palgrave Macmillan, pp. 77–97.
Stalnaker, R. C. (1989), "On What's in the Head", *Philosophical Perspectives* 3: 287–319.
Tolman, E. C. (1948), "Cognitive Maps in Rats and Men", *Psychological Review* 55: 189–208.
Tolman, E. and Honzik, C. (1930), "'Insight' in Rats", *University of California Publications in Psychology* 4: 215–232.
Vosgerau, G. (2009), *Mental Representation and Self-Consciousness. From Basic Self-Representation to Self-Related Cognition*, Paderborn: Mentis.
Vosgerau, G. (2010), "Memory and Content", *Consciousness and Cognition* 19: 838 – 846.
Vosgerau, G., Schlicht, T. and Newen, A. (2008), "Orthogonality of Phenomenality and Content', *American Philosophical Quarterly* 45: 309–328.
Vosgerau, G. and Synofzik, M. (2010), "A Cognitive Theory of Thoughts", *American Philosophical Quarterly* 47: 205–222.
Walter, S. (2010), "Cognitive extension: the parity argument, functionalism, and the mark of the cognitive", *Synthese* 177: 285–300.

Gotfrid Fosgerau

Posrednici, sadržaji i supervencija

Apstrakt

U ovom radu predstavljam argument u prilog pretpostavci da sadržaji supervenišu na posrednicima koji je zasnovan na eksplanatornoj ulozi predstava u kognitivnim naukama. Potom pokazujem da teza o superventnosti, zajedno sa eksplanatornom ulogom, implicira da su kriterijumi za individuaciju sadržaja i posrednika blisko povezani na takav način da je internalizam (eksternalizam) o sadržaju zapravo ekvivalentan internalizmu (eksternalizmu) o posrednicima. U ostatku rada, pokazujem da neki od različitih pristupa u debati proističu iz različitih istraživačkih pitanja, tačnije, pokazujem da neki od pristupa proističu iz pitanja koje se tiče uslova sticanja mentalnih predstava a drugi iz pitanja koje se tiče uslova za razmišljanje o mentalnim predstavama. Najzad, pokazujem da je teza o eksternalizmu daleko zanimljivija ako je razumemo kao tezu o tome na koji način mentalno predstavljanje funkcioniše u našem svetu a ne kao tezu o tome na koji način mentalno predstavljanje funkcioniše u svim metafizički mogućim svetovima. Tačnije, pokazujem da „nomološko“ razumevanje teze može da objasni kako i zašto eksperimentalne metode koje se koriste u savremenim kognitivnim naukama mogu da pruže uvid u obrazovanje ponašanja.

Ključne reči: eksternalizam, semantički eksternalizam, eksternalizam o posrednicima, supervencija, empirijske metode