Dialogue (2024), **63**, 575–586
doi:10.1017/S0012217323000136

ARTICLE

Success Semantics, Reinforcing Satisfaction, and Sensory Inclinations

Howard Nye[1] and Meysam Shojaeenejad[2]

[1]Department of Philosophy, University of Alberta, Edmonton, AB, Canada and [2]Independent Scholar
Corresponding author: Howard Nye; Email: hnye@ualberta.ca

Abstract

Success semantics holds, roughly, that what it is for a state of an agent to be a belief that *P* is for it to be disposed to combine with her desires to cause behaviour that would fulfill those desires if *P*. J. T. Whyte supplements this with an account of the contents of an agent's "basic desires" to provide an attractive naturalistic theory of mental content. We argue that Whyte's strategy can avoid the objections raised against it by restricting "basic desires" to sensory inclinations that cause us to do things independently of our beliefs about their contents.

Résumé

La sémantique du succès soutient, en gros, que ce qu'il faut pour qu'un état d'un agent soit une croyance en *P*, c'est que cet état soit disposé à se combiner avec ses désirs pour provoquer un comportement qui répondrait à ces désirs si *P*. J. T. Whyte complète cela par un compte-rendu du contenu des « désirs de base » d'un agent pour fournir une théorie naturaliste attrayante du contenu mental. Nous soutenons que la stratégie de Whyte peut éviter les objections qui lui sont faites en restreignant les « désirs de base » aux inclinations sensorielles qui nous font agir indépendamment de nos croyances sur leur contenu.

Keywords: success semantics; naturalistic theories of mental content; functionalism; beliefs; desires; inclinations; perceptions

1. Whyte's Success Semantics

Most generally and roughly, success semantics holds that

> **(R)** What it is for a state of an agent *B* to be a belief that *P* is for *B* to be disposed to combine with the agent's desires to cause behaviour that would fulfill those desires if *P*.

For example, what makes it the case that some (e.g., neural) state of Mary is a belief that the food in front of her is poisoned is that this state is disposed to combine with

https://doi.org/10.1017/S0012217323000136 Published online by Cambridge University Press

her desires to cause behaviour that would fulfill those desires if it was poisoned. The state would, for instance, combine with Mary's desire to eat such food only if it is not poisoned to cause her to avoid eating it.

Functionalist accounts seek to explain the nature of mental states in terms of their causal dispositions. Success semantics is a functionalist account not only of what it is for something to be an instance of a general kind of mental state — like *belief* — but what it is for it to have a particular intentional content — or be a belief *that P*.[1] Success semantics thus has the general virtues of functionalism of explaining how mental states can be multiply realized by different sorts of physical states without being something "over and above" the physical.[2] For instance, Mike's belief that sharks are present might be realized by the dispositions of certain of his cortical neurons, while beliefs with the same content in Joe, a seabird, an octopus, an alien observer, and an AI system of the future might be realized respectively by slightly different neurons in Joe's cortex, the bird's homologous but more nuclear cerebrum, the octopus's ganglia, the alien's fluid sacks, and the AI's silicon hardware.[3] R plausibly captures the key causal dispositions common to all these physical states in virtue of which we should see them as realizing beliefs with the same or at least relevantly similar content.[4]

While Frank Ramsey (1990) is typically credited with this basic idea of success semantics, J. T. Whyte (1990, 1991) has developed it into a comprehensive naturalistic theory of mental states' intentionality or content in non-intentional terms. Since R explains beliefs' contents in terms of the contents of desires, such a theory needs a naturalistic account of desires' contents. As Whyte observes, a plausible parallel functionalist account of desires' contents in terms of their causal dispositions is

> **(F)** What it is for a state of an agent *D* to be a desire that *O* (i.e., that will be fulfilled by outcome *O*) is for *D* to be disposed to combine with the agent's beliefs to cause behaviour that would bring it about that *O* if those beliefs were true (i.e., if their contents were to obtain).

For instance, the causal essence of a desire to eat food of a given kind only if it is not poisoned seems to be a disposition to combine with one's beliefs (e.g., that the food of

[1] Cf. Hattiangadi (2007, pp. 120–121). Authors who give functionalist accounts of what it is for a contentful mental state to be a belief or desire (with content tokened in the "belief-box" or "desire-box") without giving such a functionalist account of its content include Fodor (1987). For criticism of Fodor's atomistic asymmetric dependence theory of content, see Adams and Aizawa (1994, 2017).

[2] See, e.g., Braddon-Mitchell and Jackson (2007, pp. 48–49). Some functionalists are "realizer functionalists," who identify mental states with the various physical states that realize or play the causal roles themselves (e.g., Armstrong, 1981; Lewis, 1966). Other functionalists are "role functionalists," who identify mental states with the higher-order state of having some physical realizer or other that plays the causal roles (e.g., Fodor (1974).

[3] See, e.g., Edelman and Seth (2009); Jarvis et al. (2005); Mather (2008).

[4] If the content of a belief is determined by the full range of states with which it can combine to produce behaviour, then, as on a holistic version of conceptual role semantics, its content may be somewhat different from individual to individual, and sameness of content will be a matter of degree (Fodor, 1987, pp. 55–95). Some authors also suggest that there are elements of narrow content (or the way an individual conceives of what her states represent) that should be explained by something more like conceptual role semantics, although success semantics may still adequately explain broad content (or what it is the individual's states represent) — see, e.g., Bermudez (2003, pp. 88–108); Block (1986).

https://doi.org/10.1017/S0012217323000136 Published online by Cambridge University Press

that kind in front of one is poisoned) to cause one to do things (like avoid eating the food) that would bring it about that one eats food of that kind only if it is not poisoned if those beliefs were true (and e.g., the food in front of one is poisoned).

Unfortunately, as Whyte notes, a theory of content comprised of nothing but R and F seems viciously circular. The problem is not simply that R and F describe beliefs and desires as interlocking dispositional states, but rather that too little has been said about these interlocking dispositions in non-intentional terms for them to constitute a substantive naturalistic theory of content. One can use the Ramsey-Carnap-Lewis method to give a non-circular account of entities described by interlocking dispositions by replacing the names of these entities with bound variables and saying that there exists a set of entities that relate to each other in the relevant ways (Braddon-Mitchell & Jackson, 2007; Lewis, 1970).[5] But if one tries to use this approach to explain the contents of beliefs and desires using only R and F, all one can say is that there are two kinds of states which are disposed to combine to produce behaviour that fulfills or brings about the content of one if the other is true or such that its content obtains. But this does not seem to say enough to explain the properties of a state's being fulfilled or such that its content is brought about, or its being true or such that its content obtains.

[5] To illustrate this method, suppose that we want to explain what it is for an entity to be in a state described in terms of a theory. In the case of the theory composed simply of R and F, we might want to explain (without loss of generality) what it is for an agent to have the belief B_1 that p_1 (where $p_1 =$ the food in front of one has been poisoned). First, we take everything that the theory says about the state we want to explain, including everything the theory says about other states that interact with or are otherwise related to the state in question. In the case of the theory composed of R and F, this would include that B_1 is disposed to combine with desire D_1 that o_1 (where $o_1 =$ one eats the food in front of one only if it is not poisoned) to cause behaviour that would bring about o_1 if p_1 (including avoiding eating the food). It would include further that B_1 is disposed to combine with other desires to cause other behaviours — e.g., it is disposed to combine with the possible (self-destructive) desire D_2 that o_2 (where $o_2 =$ one's being poisoned) to cause behaviour that would bring about o_2 if p_1 (including one's eating the food). It would also include that D_1 is disposed to combine with other beliefs to cause other behaviours (and similarly for D_2 and all other desires with which B_1 could combine), including that D_1 is disposed to combine with belief B_2 that p_2 (where $p_2 =$ the food in front of one is *not* poisoned) to cause behaviour that would bring about o_1 if p_2 (including one's eating the food), and so on.

Next, we form what is known as the "Ramsey sentence" of the theory, which conjoins everything that the theory says about the states the meaning of which we take to be explained in terms of the theory, replaces the names of these states with variables, and says that there exist states that play these roles. In the case of R and F, the states explained in terms of the theory are the beliefs B_1, B_2, … and the desires D_1, D_2 , … So we obtain:

> The Ramsey sentence of R and F: $(\exists x_1)$ $(\exists x_2)$ … $(\exists y_1)$ $(\exists y_2)$ … [x_1 is disposed to combine with y_1 to cause behaviour that would bring about o_1 if p_1, x_1 is disposed to combine with y_2 to cause behaviour that would bring about o_2 if p_1, … x_2 is disposed to combine with y_1 to cause behaviour that would bring about o_1 if p_2, …]

Finally, we identify an entity's being in a particular state described by the theory as (i) the holding of the Ramsey sentence, which says that there are states that play all of the roles posited by the theory, and (ii) the entity's being in the state that plays the particular role described by the theory. Here, we would explain what it is for an agent to have the belief B_1 with (i) the Ramsey sentence of R and F, and (ii) the agent is in x_1.

https://doi.org/10.1017/S0012217323000136 Published online by Cambridge University Press

Still, as Whyte suggests, if we can give a naturalistic account of the contents of *some* desires that does not refer to the contents of beliefs, we can then use R and F to build upon this foundation to give an account of the contents of beliefs and other desires. Whyte's proposal is in essence that there is a set of "basic desires," the contents of which we can account for naturalistically in terms of what *reinforcingly satisfies* them:

> **(S)** What it is for a state of an agent *D* to be a basic desire that *O* is for it to be the case that *O* would (i) cause *D* to "go away" or cease exerting causal influence on the agent's behaviour, (ii) in a way that would reinforce the agent's disposition to act in the way that led to *O* when *D* is next present or active.[6]

For instance, Whyte thinks that desires like those to possess and eat cherries are basic desires: possessing cherries causes one's desire to possess them to cease influencing one's conduct, and eating cherries causes one's desire to eat them to cease influencing one's conduct. Of course, desires can cease influencing one's conduct through the obtaining of things other than their content — such as one's receiving a hard blow to the stomach. But when one lacks a desire for a blow to the stomach, one's receiving a blow does not tend to cause one to repeat whatever led to one's receiving it. By contrast, it is plausibly essential to basic desires or inclinations like those to possess or eat cherries that one has a tendency to become more likely to do whatever it was that led to one's possessing or eating them if one is motivated by those desires in the future.[7]

2. Problems for the Reinforcing Satisfaction Account

Whyte thinks, however, that S as it stands fails to provide either necessary or sufficient conditions for an outcome to be the content of, or what fulfills, a basic desire. He claims that an outcome *O* can be the content of a basic desire *D*, even though it is *not* the case that *O* would (i) cause *D* to go away (ii) in a way that

[6] And, if we are to understand this as a full account of what it is for a state to be a basic desire with a certain content (instead of just what it is for a state assumed to be a basic desire to have a certain content), we should add something like "and (iii) *D* is disposed to combine with the agent's beliefs to produce behaviour in the way described by R." Because Whyte explicitly states R, F, and S as accounts only of states' contents or truth and fulfillment conditions, he does not add any such clause. Because we think that it is much clearer to give R, F, and S as accounts of what it is for a state to be a belief or desire with a certain content (as we do in the text), our way of stating S requires this third clause. But because it is natural to assume that the states described by S's (i) and (ii) also have feature (iii), we omit (iii) from the discussion in the text for simplicity.

[7] Note that this account is dispositional or counterfactual in several ways: one can have a basic desire that *O*, which is never actually reinforcingly satisfied, and if it is reinforcingly satisfied, one might never again come to be moved by the desire or have the opportunity to do whatever led to its previous satisfaction. The account requires only that *O would* reinforcingly satisfy *D if O were to* occur, and *D*'s being reinforcingly satisfied requires only that one *would* have a *tendency* to do what led to *O* in the past *if one were* to be moved by *D* again in the future. Note also that these dispositions can be masked by those associated with similar basic desires. As Whyte (1991, p. 67) notes, one can have a basic desire to eat cherries, eat a few cherries, and still have a basic desire to eat cherries that has not gone away. But, in this case, one has several basic desires for several cherries, some but not all of which have been satisfied (or a basic desire for a certain quantity of cherries, which has not yet been satisfied).

https://doi.org/10.1017/S0012217323000136 Published online by Cambridge University Press

reinforces the behaviour that led to *O*. For instance, Whyte suggests that Prema's desire to possess cherries could be fulfilled by someone's putting cherries into her handbag when she is not looking. Especially if Prema never comes to know when or how she got the cherries (or even that she has them), it seems that this fulfillment of her desire to possess cherries could fail to reinforce any of the behaviours (e.g., her approaching the individual who surreptitiously gave them to her) that led to its fulfillment.[8]

Whyte also claims that it is possible for outcome *O* to cause basic desire *D* to (i) go away (ii) in a way that reinforces the behaviour that led to *O*, without *O* actually being the content of *D*. For instance, suppose that Prema cannot distinguish genuine cherries from similar-tasting imitation cherries, but that she nonetheless desires to eat only genuine cherries, say because she believes that they are healthier. Whyte suggests that Prema's eating imitation cherries could (i) cause her desire to eat genuine cherries to go away, (ii) in a way that reinforces her tendency to do whatever led to her eating imitation cherries — such as her ordering "cherries" from someone she thinks is selling her genuine ones — if she is moved by her desire for genuine cherries in the future.

In response to these problems, Whyte (1991, pp. 70–73) proposes a modification of S, the basic idea of which is that a basic desire's content is what would reinforcingly satisfy it under "normal conditions." Thus, a state of Prema's can be a basic desire to possess cherries despite its not being reinforcingly satisfied by possessing cherries she does not notice, and a state of Prema's can be a basic desire to eat genuine cherries despite its being reinforcingly satisfied by imitation cherries, because her being unable to distinguish cherries' absence from presence and genuineness from imitation is relevantly "abnormal." Whyte notes, however, that it would be viciously circular to understand "normal" conditions as whatever conditions are such that under them a basic desire is reinforcingly satisfied just in case it is fulfilled. He also notes that it would be a mistake to identify normal conditions with whatever conditions are statistically typical, since it is possible for error to be statistically typical. Whyte proposes to understand normal conditions as those that would remain reinforcingly satisfying no matter how much the agent's perceptual capacities were improved, or

> **(S')**: What it is for a state of an agent *D* to be a basic desire that *O* is for it to be the case that *O* would (i) cause *D* to go away (ii) in a way that would reinforce the agent's disposition to act in the way that led to *O* when *D* is next active, and (iii) (i) and (ii) would remain true no matter how much the agent's perceptual capacities were improved.

[8] We acknowledge that it might at first seem strange to have a desire to possess cherries that could be fulfilled even if one never comes to know that it is fulfilled. We think that such desires are not centrally important to this example (which we develop in parallel to Whyte's discussion); all that is really important for our purposes are cases of a desire to possess cherries that is not reinforcingly satisfied by the behaviours that lead to its fulfillment because one does not realize how one came to possess the cherries. These include cases where the agent's coming to realize that she has the cherries is important to her desire being fulfilled, in virtue of her desiring to possess cherries only if she later eats them. But that said, we think that desires of an agent to possess cherries *per se*, which can be fulfilled without the agent ever knowing it, might not be so odd if, for example, they are part of the agent's allocating scarce deliberative resources over time by focusing only on coming to possess cherries while searching for them, without remembering why she has formed the desire or what broader goals it is serving. We are grateful to an anonymous reviewer for *Dialogue* for pressing us on this point.

https://doi.org/10.1017/S0012217323000136 Published online by Cambridge University Press

Anandi Hattiangadi has objected that S' is viciously circular. As she puts it, "In order to decide what counts as an 'improvement' of my perceptual abilities, assumptions have to be made about what I want, which is ultimately circular" (Hattiangadi, 2007, p. 125). We do not think, however, that the circularity is as obvious as Hattiangadi appears to suppose. It might seem that Whyte could characterize perceptual improvements as something like increased abilities to discriminate among or respond differentially to different states of the world. Such an account of improvement would not obviously presuppose the content of the agent's desires.

That said, we think Hattiangadi is correct that the plausibility of S' trades upon a kind of vicious circularity. It does not seem that the contents of an agent's desires should depend upon what would reinforcingly satisfy them if she underwent any arbitrary increase in discriminative ability. For instance, what would be true in fanciful scenarios such as Prema's gaining the ability to perceptually discriminate among subatomic particles does not seem relevant to the content of her desires like those to possess and eat cherries. For S' to be a plausible explanation of the content of Prema's desires to possess and eat cherries, it needs instead to consider such things as what would be true if she were able to discriminate between present and absent cherries and between genuine and imitation cherries. But to distinguish these discriminations as relevant from others that are not seems to presuppose the contents of Prema's desires in a viciously circular way.

Moreover, an arbitrary increase in an agent's discriminative abilities might alter the contents of her desires. This would cause (i) and (ii) to cease to hold for some outcome *O* not because *O* was not what the agent desired, but because the changes mentioned in (iii) change what she desires. For example, Prema might currently have a desire to eat cherries that would be fulfilled by cherries with a wide range of acidity levels, but improving her range of gustatory discrimination would cause her to desire to eat only cherries with a very particular acidity level. It does not seem that Whyte's approach to solving the problems he raises with S allows him to distinguish between perceptual improvements that do as opposed to do not alter basic desires' contents without viciously circular assumptions about what their contents actually are.

3. Basic Desires as Sensory Inclinations

We think, however, that there is an alternative version of S that avoids the problems that Whyte raises for it, which avoids appealing to "normal conditions" and thus also avoids the problems with S'. Our solution looks more carefully at the role that should be played by Whyte's "basic desires." These should be simple motivational states of an agent, the causal tendencies of which do not depend upon her beliefs about their contents. But Whyte's examples of allegedly "basic" desires that make trouble for S are actually desires like those *to possess* cherries and to eat *genuine* as opposed to imitation cherries.[9] Whether these desires are reinforcingly satisfied obviously depends upon whether the agent believes that their contents obtain.

[9] Whyte's (1991, p. 70) own examples are of a "desire for chocolate" — which he suggests can be fulfilled but not reinforcingly satisfied by someone's putting chocolate in one's trouser pocket, and reinforcingly satisfied but not fulfilled by one's eating carob that one cannot distinguish from chocolate.

https://doi.org/10.1017/S0012217323000136 Published online by Cambridge University Press

But agents' most phylogenetically and ontogenetically conserved motivational states do not seem to be like this. These motives include the urges, likes, and dislikes involved in (or at least accompanying) such sensory states as itches, hunger, thirst, pleasurable experiences, and affectively painful experiences.[10] The objects of these *sensory inclinations* are internal states of the agent herself, such as her experiencing scratching, experiencing the ingestion of food or water, or her continuing or ceasing to have certain experiences. These motives are reinforcingly satisfied by the obtaining of the internal states of the agent that are their content, rather than the agent's beliefs or perceptions that these states obtain.[11]

To the extent that sensory inclinations are motives for the agent to have certain subjective experiences, it is arguable that her introspective representations of whether they are fulfilled cannot be false. But it is also arguable that these experiences can occur in the absence of beliefs or representations about them, and when they do, they reinforcingly satisfy sensory inclinations for them. Many agents, such as human infants and some sentient non-human animals, arguably have sensory inclinations that are reinforcingly satisfied by the experiences that are their contents, although they lack the capacity to form representations to themselves about whether they are having these experiences. When we potentially reflective agents are too busy experiencing the world and acting upon it to indulge in navel-gazing reflection about

[10] See, e.g., Hall (2008). Hall suggests that we should distinguish these motivational states from desires, because they stand to desires as perceptions stand to beliefs. Perceptions and these motivational inclinations are "peripheral" in that they set default behavioural tendencies, while relaying information and commands to the "central" or domain general system of beliefs and desires, which are informed by but can override the behavioural tendencies of perceptions and inclinations (Hall, 2008, pp. 532–533). Kahane (2009) similarly argues that the kind of dislike involved in suffering the sort of pain that is intrinsically bad should be distinguished from bare desires, because this dislike is necessarily experienced. Whether motivations and representations the contents of which are arguably the most explanatorily basic (such as inclinations and perceptions) must be subjectively or phenomenally experienced is beyond the scope of this article (as is what kind of content and other features a state must have in order to be phenomenally experienced). Our primary intention is simply to emphasize the analogy of the relationship between inclinations and ordinary desires to that between perceptions and ordinary beliefs.

[11] While we suspect that the relevant internal states may always be experiences, we are here leaving it open that they could be internal states that fall short of being experiences. We are, however, working with a notion of an internal state of an agent that is relevantly "intrinsic" to her, in the sense that the state occurs inside the agent and could in principle continue to occur if the agent's environment were radically altered. Thus, the agent's knowingly becoming the King of France would not be an internal state of the agent (as it depends upon the agent's relationship to actually becoming King and reliably tracking the fact that he is), and could not qualify as the content of a sensory inclination. Similarly, an agent who has exquisite taste in wine could have as an internal state, and content of a sensory inclination, the exact qualitative experience that in practice can only be derived from drinking a 1999 Chateau Lafite. But the actual fact that the agent is drinking a 1999 Chateau Lafite is too external and relational to count as an internal state of the agent in our sense. Because in our view the distinctive qualitative experience of drinking a 1999 Chateau Lafite must be realized by some kind of internal (e.g., a neural) state within an agent, we think that it could in principle (but not in practice) be brought about in the absence of the agent's actually engaging in the external, physical, relational act of drinking a 1999 Chateau Lafite — e.g., through direct neural stimulation that induces the relevant state, say if the agent's brain had been removed and was interacting withthe simulation program that we discuss below. We are grateful to an anonymous referee for *Dialogue* for pressing us on this point.

https://doi.org/10.1017/S0012217323000136 Published online by Cambridge University Press

how we feel, we are arguably identical in this respect to our younger and specifically different fellows.[12]

For instance, an infant, vole, or adult human who lacks time to reflect on what she is experiencing might like and have a sensory inclination to experience cherry-like taste. Having an experience of cherry-like taste will cause this inclination to (i) go away (ii) in a way that will tend to make her more likely in the future to do whatever caused her to have this experience the next time she is moved by the inclination. This can all be true without her ever having a tendency to represent to herself that she is experiencing cherry-like taste. Or consider a human adult who has a sensory inclination to experience cherry-like taste that she mistakes for a motivation to eat genuine as opposed to imitation cherries. This adult might discover that she is mistaken about her own motives by seeing how her cravings are reinforcingly satisfied by imitation just as much as by genuine cherries.

It seems then that we should offer S as an account of the content of sensory inclinations, which *are* reinforcingly satisfied by the states of the agent that are their content, rather than the content of Whyte's "basic desires" which are reinforcingly satisfied by the agent's beliefs that their contents obtain. We thus arrive at

> **(S*)** What it is for a state of an agent *SI* to be a sensory inclination that internal state *E* of the agent obtains is for it to be the case that *E* would (i) cause *SI* to go away (ii) in a way that would reinforce the agent's disposition to act in the way that led to *E* when *SI* is next active.

4. Conclusion: A Foundation for Mental Content

With S* in place, we can use R to give an account of what it is for an initial set of "level 1" representational states to have the content that *P* in terms of their tendency to combine with the agent's sensory inclinations to cause behaviour that would fulfill her inclinations if *P*. As the most basic representations with which sensory inclinations combine to influence behaviour, these might actually be more plausibly understood as perceptions or sensory representations than beliefs (see Hall, 2008, pp. 532–533). Having done this, we can then use F to give an account of what it is for "level 2 motivations" to have the content that *O* in terms of their tendency to combine with level 1 representations to cause behaviour that would bring about *O* if the content of the level 1 representations were to obtain. From here we can iterate further, giving an account of what it is for "level 2 representations" to have the content that *P* in terms of their tendency to combine with level 1

[12] See, e.g., Bargh and Chartrand (1999); Lakoff and Johnson (1999); Smith (2017, §5). Authors like Gennaro (2004) contend that infants and all sentient non-human animals can conceive of their experiences, but a key worry is that such contentions lack independent support beyond our reasons to think that they are simply having experiences as opposed to conceiving of them (and the higher-order-thought theory that Genarro uses to connect experiences to consciousness of them may not, for independent reasons, be the best theory of phenomenal consciousness). But even if Genarro were correct, he argues convincingly that it should not be possible for consciousness conferring beliefs about our experiences to misrepresent them, which would prevent our account of the content of sensory inclinations from facing Whyte's worries about reinforcing satisfaction coming apart from fulfillment due to deception.

https://doi.org/10.1017/S0012217323000136 Published online by Cambridge University Press

motivations and level 2 motivations to cause behaviour that would fulfill them if *P*, and so on.[13]

What, however, if the states of an agent that are the objects of her sensory inclinations are experiences that themselves have representational contents, so that, for instance, a painful sensation she is inclined to terminate represents tissue damage at a particular bodily location (cf. Tye, 1995)? One possibility is that the contents of sensory or perceptual representations should be explained by a different theory, such as a causal covariation account (Tye, 1995, pp. 100–105), and using this together with S*'s account of the contents of sensory inclinations as a foundation, success semantics should seek to use R and F to explain only the contents of beliefs and desires proper.

Alternatively, we do not think that it would be viciously circular to use S*, R, and F to offer a fully general theory of mental content, including the representations of the objects of sensory inclinations. On this theory, the dispositions described by R and S* are interlocking, but we think that S* can still provide enough independent traction on the idea of what it is for a sensory inclination to be fulfilled or unfulfilled to make substantive sense of the behavioural dispositions that R appeals to in explaining the contents of representations that combine with these motives. This in turn allows for a substantive understanding of the dispositions that F appeals to in explaining the contents of motives that combine with these representations, and so on. For instance, S* explains what it is for an inclination of an agent to be fulfilled by the termination of her experience of tissue damage in terms of this termination reinforcingly satisfying the inclination. This allows us to explain what it is for the experience of tissue damage itself to involve a level 1 representation of tissue damage in terms of its combining with the inclination and other representations that would cause behaviour that would fulfill the inclination if these representations were accurate. For instance, the experience's representation of tissue damage would combine with the inclination to terminate the experience and the representation that grasping the limb where the tissue damage seems to be occurring will terminate the experience to cause behaviour like reaching to grasp the limb that would fulfill (i.e., reinforcingly satisfy) the inclination if the tissue damage is in the limb and grasping it will terminate the

[13] The full iterative account is spelled out in Shojaeenejad (2017) (although there all motivational and representational states are referred to respectively as "desires" and "beliefs"). To clarify, the main dependence of higher-level representational and motivational states on lower-level ones is metaphysical rather than necessarily genetic; what it is for an agent to have the higher-level states is for the agent to have states that are disposed to interact in various ways with other states of the same level and lower levels. We do suspect that in practice the higher-level states (such as those about abstract objects) of all naturally existing agents will also genetically depend upon the agents first having lower-level states and then developing higher-level states on their basis. But we do not think that this is necessary in principle. We think that it is metaphysically possible for an agent to come into existence spontaneously with both lower-level and higher-level mental states. We believe that one such agent would be Davidson's (1987) Swampman, who, as the result of an amazing coincidence spontaneously comes into existence in a swamp, and happens to be qualitatively identical to Davidson down to the molecular level, but with no causal role played by Davidson's existence in Swampman's formation or composition. We think that Swampman would have mental states much like those of Davidson, including of the same higher- and lower-levels (although we do think that their exact broad content, or what if anything they refer to and are made correct or incorrect by, would differ in some respects). We are grateful to an anonymous reviewer for *Dialogue* for encouraging us to clarify this point.

https://doi.org/10.1017/S0012217323000136 Published online by Cambridge University Press

experience (and this is not, for example, a case of referred or phantom limb pain in which reaching for the limb will not terminate the pain). The fact that the experience which reinforcingly satisfies the inclination itself has a representational content of tissue damage occurring in a location and needs to be explained in this way does not seem to interfere with the cogency of the explanation of either the inclination's motivational or the experience's representational content.

While the content of sensory inclinations is restricted to internal states of the agent, the content of level 1 representations will include the conditions of the external world that guarantee (or make more likely[14]) the fulfillment of her sensory inclinations by the behaviours motivated by these representations and inclinations. An agent's level 1 representations might represent that there is an object in front of her; it looks cherry-like (it is small, round, red, etc.); if it looks cherry-like then it will taste cherry-like; the cherry-looking object is not hard; if the object is hard then chewing it will be painful; etc. The agent could be wrong about any of this: she could be hallucinating the object, expecting an olive that feels cherry-like to look cherry-like, expecting a cherry-looking cherry tomato to taste cherry-like, expecting a cherry-looking painted rock not to be hard, etc. If so, then her representations and inclinations will combine to cause behaviour that would fulfill the inclinations if these conditions obtained, but will, for some inclinations and eventualities, fail to fulfill them.

Since the contents of an agent's level 1 representations are the conditions under which these behaviours yield outcomes that reinforcingly satisfy her sensory inclinations, her level 1 representations are made correct or incorrect by the external world in a familiar sense in standard environments. But level 1 representations make few commitments about the underlying nature of the world, as is plausible for basic perceptions or perceptual beliefs. Suppose that the agent's brain were removed from her body, placed in a nutrient vat, and connected to a stimulation program that involves the exact same contingencies between her behaviour (understood as motor command output) and those sensory states of her brain that reinforcingly satisfy her sensory inclinations. Her level 1 representations would now be made correct or incorrect by features of the program instead of facts about the interaction of her body with other medium-sized objects.[15]

An agent must ascend farther up the hierarchy of iterations of F and R in order to represent or care about the difference between entities that are fully experientially indistinguishable, like whether she is eating genuine or imitation cherries, or whether the local watery stuff is H_20 or XYZ. To do this, she must along the way develop something like representations or motivations that concern whether certain things are caused by other things, and whether various of her experiences have similar

[14] Whether success semantics is best formulated in terms of truth guaranteeing success (Dokic & Engel, 2002; Whyte, 1990) or increasing the probability of success (Nanay, 2013) is a matter of some controversy. For further discussion, see Shojaeenejad (2017).

[15] For instance, the sense in which her level 1 representations represent something as "an object in front of her" would be sufficiently neutral as to be made correct or incorrect by her body's relation to other physical objects before her brain is removed and connected to the program, as well as made true by those states of the program that give rise to the same causal relationships between motor outputs and experiences after her brain is removed and connected to the program.

https://doi.org/10.1017/S0012217323000136 Published online by Cambridge University Press

causes (see, e.g., Chalmers, 2012, pp. 312–378). We leave the investigation of what success semantics should say about how exactly this works to another occasion. But we hope that we have said enough here to make clear how, by using reinforcing satisfaction to explain the content of an agent's inclinations for her own sensations, success semantics can, as Whyte hoped, successfully lay a foundation for explaining the content of her other mental states.

Competing interests. The authors declare none.

References

Adams, F., & Aizawa, K. (1994). Fodorian semantics. In S. Stich & T. Warfield (Eds.), *Mental representation: A reader* (pp. 223–242). Blackwell. https://www.wiley.com/en-us/Mental+Representation%3A+A+Reader-p-9781557864772

Adams, F., & Aizawa, K. (2017). Causal theories of mental content. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2017 Edition). Stanford University. https://plato.stanford.edu/archives/sum2017/entries/content-causal/

Armstrong, D. M. (1981). The causal theory of the mind. In D. M. Armstrong (Ed.), *The nature of mind and other essays* (pp. 16–31). University of Queensland Press.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*(7), 462–479. https://doi.org/10.1037/0003-066X.54.7.462" https://doi.org/10.1037/0003-066X.54.7.462

Bermudez, J. L. (2003). *Thinking without words*. Oxford University Press. https://global.oup.com/academic/product/thinking-without-words-9780195341607?cc=us&lang=en&

Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, *10*(1): 615–678. https://onlinelibrary.wiley.com/doi/10.1111/j.1475-4975.1987.tb00558.x

Braddon-Mitchell, D., & Jackson, F. (2007). *Philosophy of mind and cognition: An introduction* (2nd ed.). Wiley-Blackwell. https://www.wiley.com/en-us/Philosophy+of+Mind+and+Cognition:+An+Introduction,+2nd+Edition-p-9781405133234

Chalmers, D. (2012). *Constructing the world*. Oxford University Press. https://global.oup.com/academic/product/constructing-the-world-9780199608584?cc=us&lang=en#&

Davidson, D. (1987). Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association*, *60*(3), 441–458. https://doi.org/10.2307/3131782

Dokic, J., & Engel, P. (2002). *Frank Ramsey: Truth and success*. Routledge. https://www.routledge.com/Frank-Ramsey-Truth-and-Success/Dokic-Engel/p/book/9780415408288

Edelman, D. B., & Seth, A. K. (2009). Animal consciousness: A synthetic approach. *Trends in Neurosciences*, *32*(9), 476–484. https://doi.org/10.1016/j.tins.2009.05.008

Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese*, *28*(2), 97-115. https://doi.org/10.1007/BF00485230

Fodor, J. A. (1987). *Psychosemantics*. MIT Press. https://mitpress.mit.edu/9780262061063/psychosemantics/

Gennaro, R. J. (2004). Higher-order thoughts, animal consciousness, and misrepresentation: A reply to Carruthers and Levine. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness: An anthology* (pp. 45–66). John Benjamins. https://doi.org/10.1075/aicr.56

Hall, R. J. (2008). If it itches, scratch! *Australasian Journal of Philosophy*, *86*(4), 525–535. https://doi.org/10.1080/00048400802346813

Hattiangadi, A. (2007). *Oughts and thoughts: Scepticism and the normativity of meaning*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199219025.001.0001

Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., Medina, L., Paxinos, G., Perkel, D. J., Shimizu, T., Striedter, G., Wild, J. M., Ball, G. F., Dugas-Ford, J., Durand, S. E., Hough, G. E., Husband, S., Kubikova, L., Lee, D. W., … Butler, A. B. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience*, *6*(2), 151–159. https://doi.org/10.1038/nrn1606

Kahane, G. (2009). Pain, dislike and experience. *Utilitas*, *21*(3), 327–336. https://doi.org/10.1017/S0953820809990070

https://doi.org/10.1017/S0012217323000136 Published online by Cambridge University Press

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. Basic Books. https://www.hachettebookgroup.com/titles/george-lakoff/philosophy-in-the-flesh/9780465056743/?lens=basic-books

Lewis, D. K. (1966). An argument for the identity theory. *Journal of Philosophy*, *63*(1), 17–25. https://doi.org/10.2307/2024524

Lewis, D. K. (1970). How to define theoretical terms. *Journal of Philosophy*, *67*(13), 427–446. https://doi.org/10.2307/2023861

Mather, J. A. (2008). Cephalopod consciousness: Behavioural evidence. *Consciousness and Cognition*, *17*(1), 37–48. https://doi.org/10.1016/j.concog.2006.11.006

Nanay, B. (2013). Success semantics: The sequel. *Philosophical Studies*, *165*(1), 151–165. 10.1007/s11098-012-9922-7

Ramsey, F. P. (1990). Facts and propositions. In D. H. Mellor (Ed.) *F. P. Ramsey: Philosophical papers*. Cambridge University Press. (Original work published 1927.) https://www.cambridge.org/us/academic/subjects/philosophy/philosophy-texts/f-p-ramsey-philosophical-papers?format=PB

Shojaeenejad, M. (2017). *Success semantics: Motivations, problems, and solutions* [Master's thesis, University of Alberta]. Edmonton. https://doi.org/10.7939/R3KH0FC70

Smith, J. (2017). Self-consciousness. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2017 Edition). Stanford University. https://plato.stanford.edu/archives/fall2017/entries/self-consciousness/

Tye, M. (1995). *Ten problems of consciousness: A representational theory of the phenomenal mind*. MIT Press. https://doi.org/10.7551/mitpress/6712.001.0001

Whyte, J. T. (1990). Success semantics. *Analysis*, *50*(3): 149–157. https://doi.org/10.1093/analys/50.3.149

Whyte, J. T. (1991). The normal rewards of success. *Analysis*, *51*(2): 65–73. https://doi.org/10.1093/analys/51.2.65

Cite this article: Nye, H., Shojaeenejad, M. (2024). Success Semantics, Reinforcing Satisfaction, and Sensory Inclinations. *Dialogue* 63(3), 575–586. https://doi.org/10.1017/S0012217323000136