The Journal of Mind and Behavior
Spring 2004, Volume 25, Number 2
Pages 79–96
ISSN 0271–0137

On the Reclamation of a Certain Swampman

Mazen Maurice Guirguis

Kwantlen University College

A currently popular form of psychological externalism takes the causal–evolutionary history of a person to be determinant of that person's intentional content. Two challenges bearing on the feasibility of this doctrine are outlined and discussed: the problems of functional indeterminacy and the psychological non-status of Davidson's (1998) Swampman. Using Schank and Abelson's (1977) script construct, a division of intentionality into an aboutness component (conceived causally–evolutionarily) and a directedness component (defined with the help of the mathematical notion of an equivalence class) is introduced.

Keywords: intentionality, content, internalism

Intentionality is that property of some mental states by which these states are represented as being *directed toward*, or *about*, or *of* objects, events, or states of affairs. Hence any mental state with an intentional nature takes something *specifiable* as its object.[1] Intentional *objects* are here interpreted broadly to include living organisms, fictitious characters, inanimate items, events, states of affair, and the like. Beliefs, desires, fears, hunches, and inclinations are typical intentional states, since a belief must be a belief *that* such-and-such is (was, will be) the case; a desire must be a desire *for* someone, something, or *that* some event should come to pass; a fear must be a fear *of* something, someone, or some circumstance; and so on. On the other hand,

I owe a debt of gratitude to the reviewers of the *Journal of Mind and Behavior* for the many helpful suggestions and constructive criticisms they provided. I must also thank professors Steven Savitt, Paul Bartha, and Mohan Matthen — all members of the Philosophy Department at the University of British Columbia — for their comments on a number of ancestral drafts of this paper. Requests for reprints should be sent to Mazen M. Guirguis, Ph.D., Department of Philosophy, Kwantlen University College, 12666-72nd Avenue, Surrey, British Columbia, Canada V3W 2M8. Email: Mazen.Guirguis@kwantlen.ca

[1]The common usage of "intentional" — referring to a deliberate plan or disposition to take a certain action — should not be confused with the technical sense we are employing here.

some mental states are not so focused. "Raw feels" like pains, itches, and tickles are normally considered non-intentional.[2]

Psychological *externalists* — or *anti-individualists*, as they are sometimes known — hold that the content of intentional thought is not exclusively fixed by what is going on inside someone's skin; the agent's environment also has a role to play.[3] Just what contribution the environment makes depends on what one means by "environment." Different intuitions have yielded different arguments.

One way to think of the environment is in terms of the *objects* that surround us. In the context of externalism, this means that the physical furniture of our lives is relevant to the individuation of our intentional states. This is brought out nicely in Putnam's (1975) famous Twin-Earth thought-experiment. Another way to conceive of the environment is in terms of the social and linguistic practices of a community, for it appears that sociolinguistic conventions have an equal right to define the nature of intentional content as do sticks and stones. Burge (1998) has long adhered to this interpretation. Moreover, the environment includes the causal history or etiology of events. Here the word "history" is a variable as well, for it can signify durations ranging from the entire evolutionary development of a species to the experiential life span of a single organism. Davidson (1998) assumes a relatively local timeframe, arguing — with the aid of yet another famous thought-experiment — that the intentional content of mental events depends most importantly on the causal context in which thought is produced in an agent. In contrast, Millikan (1984) interprets "environment" more generously to include the entire evolutionary heritage of the thinker.

There is reason to approach anti-individualism in general with caution, but I want to focus on teleologically-motivated externalism in this paper, the sort that takes causal–evolutionary history to have the final word concerning what a mental state is about. I find such theories unpersuasive. They suffer from two distinct difficulties which, to my mind, render them inadequate: the first is the problem of *functional indeterminacy*; the second involves the much-maligned (and grievously underestimated) *Swampman*.

---

[2]There is no universal consensus on this point, however. See, for instance, Crane's (2003) defense of intentionalism — the view that all mental states (including raw feels) are intentional.

[3]That, at least, is how the doctrine has often been expressed. But a quick glance at the literature will reveal that anti-individualist pedagogy often renders external factors so inflated that they leave little or no room for anything else.

Causal–Evolutionary History

Most discussions of intentionality eventually make reference to Putnam's (1975) classic article, "The Meaning of 'Meaning,'" in which we are to conceive of a near duplicate of our planet Earth, called "Twin-Earth." Twin-Earth resembles Earth in nearly every detail. The physical environments look and largely are the same. Many of the inhabitants of one planet have counterparts on the other, with identical microphysical, experiential, and dispositional states and histories. There is one difference, however. It just so happens that on Twin-Earth there is no H_2O. The liquid that runs in rivers on the twin planet, that fills bathtubs and falls from the sky is perceptibly similar to H_2O, but is in fact a different compound with a very different chemical structure, XYZ. The inhabitants of Twin-Earth call XYZ "water," but twin-water ($water_{te}$)[4] is not water: *water* is H_2O. The year is 1750, when no one on Earth is yet aware of the molecular composition of water, and scientists on Twin-Earth have not yet discovered that $water_{te}$ is XYZ.

We now suppose that Adam is an English-speaking native of Earth and that $Adam_{te}$ is his physiological duplicate on Twin-Earth.[5] When Adam and his doppelgänger simultaneously form beliefs that they express by saying, "There is water in the pitcher," what they *say* is different, since their respective utterances have different truth-conditions. The twins' *beliefs*, it is contended, will also be different, since their beliefs pick out different objects in their respective environments: Adam's belief picks out water, H_2O; $Adam_{te}$'s picks out XYZ, twin-water. The conclusion we are invited to draw from this is that the physiological identity of the twins does not guarantee the identity of their intentional states.

The moral of Putnam's scenario has often been taken to be that the causal history of a person contributes essentially to what his or her intentional states *pick out* or *refer to* in the external world. The reason Adam's thoughts are about H_2O but $Adam_{te}$'s are about XYZ is allegedly due to corresponding differences in the doppelgängers' biographies: Adam's history connects him to H_2O, $Adam_{te}$'s connects him to XYZ. A theory that promotes the importance of such a historical connection, and incorporates it with Dretske's

---

[4]The subscript "te" stands for "Twin-Earth."

[5]By "physiological duplicate" I mean that we have two individuals who are *microstructurally* identical. But the twins are different at least in one respect. When they are counted, they will be assigned different numbers. That is to say that the twins occupy different points in space, and will therefore have different spatial relationships to any given external object (e.g., Earth or Twin-Moon). So at least in one sense of "physical" — a broad sense that includes external relational properties — Adam and $Adam_{te}$ are *not* physically identical. But this is irrelevant to the present point. The premise is that the twins are *physiologically* or *molecularly* indistinguishable, and this says nothing of the distal relations they might or might not have with the objects around them.

(1990) information–theoretic account of misrepresentation, has recently been advanced by Stalnaker (1999a):

> The theoretical account I have in mind is the information–theoretic account of intentional content. The rough idea is this: states of mind *carry information* when there exists a pattern of counterfactual dependencies between those states and corresponding states of the environment. If x is in a state caused by the fact that P, and would not have been in that state if it had not been that P, then that state of x carries the information that P. . . . [A]ccording to the information–theoretic picture, misrepresentation must be understood as a deviation from a norm. It is reasonable to assume that representational states are *normally* correct — that they are states that *tend* to represent things as they are. Given an appropriate conception of normal conditions, we can explain representation generally in terms of information: a state represents the world as being such that P, and so is a state with informational content P, if and only if under normal conditions it would carry the information that P. (p. 214)

Anti-individualists have often claimed that once things are described in this way, it becomes difficult to see why positing a *narrow* content — a kind of content that simultaneously captures and is limited to a first-person perspective — is necessary at all. We can, so the argument goes, fix the intentionality of a mental state externally, and account for misrepresentation by using a context-dependent notion of "normal" that brings in facts about the causal–evolutionary history of the organism.[6] Stalnaker (1999b) thus attacks Dennett's (1987) attempt to isolate an individual's "organismic contribution" to the content of his or her beliefs and other propositional attitudes. Dennett calls his idea *notional attitude psychology*, and contrasts it both with *propositional attitude psychology* (which describes attitudes in terms of the ordinary *wide* or extra-organismic, environment-dependent conception of content) and *sentential attitude psychology* (which takes the contents of attitudes to be sentences of an inner language).

The motivation behind notional attitude psychology is to explain how the purely internal properties of an individual can be used to identify a set of possible worlds compatible with that individual's narrow intentional state. Adam believes correctly that there is water in the pitcher. The proposition he believes is true on Earth where "water" rigidly designates H_2O, but false on Twin-Earth where "water" rigidly designates XYZ.[7] Dennett, however, suggests that there is another proposition — roughly, that there is some water-like stuff in the pitcher — which does not distinguish the actual world from Twin-Earth. Isolating narrow content in this way does not require reference to Adam's past; what it requires is an analysis of how Adam's brain goes about its representational work. "Our task," says Dennett (1987), "is like the

[6]Others, of course, disagree. For a spirited defense of narrow content see Segal (2000).

[7]For a discussion of rigid designation see Kripke (1980).

problem posed when we are shown some alien or antique gadget and asked: what is it for?" (p. 155).

It is hardly surprising that Stalnaker (1999b) admonishes Dennett's hypothesis as something that "doesn't look like what we want at all. Possible worlds picked out in this way look more like worlds in which the organism's needs or wants are satisfied than like worlds in which its beliefs are true" (p. 182). Stalnaker insists that beliefs are states that help the believer to cope with his or her surroundings, and that the contents of these states are essentially connected to the kind of environment they help the believer to cope with. Most would agree with this, but just what is it about trying to understand Adam's perspective that abandons this idea? The requirement that Adam's intentional states must help him navigate the world does not entail that the world-according-to-Adam must be perfectly congruent with the world-as-it-is; what it does suggest is that whatever incongruence there might be could not have been evolutionarily relevant. So the fact that Adam does not distinguish H_2O from XYZ tells us something important about the kind of information-processing system he is, but it does not undermine his ability to be guided by his "water" beliefs *because there is no (and never has been) XYZ on Earth.*

The Problem of Functional Indeterminacy

Despite Stalnaker's criticisms, I think Dennett is onto something important. His observations help to uncover one of two serious challenges facing causal–evolutionary theories of content. The following from Millikan (1993) puts the first concern in perspective:

> If I can make it plausible that the entities that folk psychology postulates are indeed defined by their proper functions, and make plausible that the proper functions with which folk psychology endows these entities very likely *are* had by some special parts or states of the body, that should be enough to show that cognitive science can probably use folk psychology as a starting point. The job of cognitive science would then be, in part, to explain what the Normal constitution of these psychological entities is and *how* they Normally perform their defining proper functions. (p. 61)

Like Stalnaker, Millikan believes that to understand intentionality we must look upon the brain as a tool that performs a *proper* function in *Normal* conditions, where "propriety" and "Normalcy" are grounded in the evolutionary development of persons. But now concerns about *functional indeterminacy* become rather pressing. An analogy might be helpful here. Suppose a mechanically inclined gift-shop owner decides to build a gift-wrapping machine in an attempt to improve customer service. He does so, and finds the machine very useful. After being loaded with wrapping paper and boxes of various sizes, the machine sets about wrapping the boxes quickly and efficiently.

Despite the obvious differences, we can think of the shopkeeper as having assumed the role of evolution: just as species evolve in response to certain selectional pressures in their environment, thereby acquiring "biologically proper" functions, we can think of the machine as having been designed in response to certain pressures in the store's environment. We can then say that the machine has been given a "mechanically proper" function — the gift-wrapping operation for which it was designed.

But now consider, the machine can perform other tasks *and with no more resources than those available to it by virtue of its original design.* For instance, the gift-wrapper can be used to prepare small parcels for the post office by wrapping them in brown paper. In this way, it can be used to arrange the mail-order catalogues for shipping to the store's off-site clientele. Or it might be used to package smaller boxes together into single units (for storage or deliveries or whatever). In light of these "newly discovered" abilities, can we still consider the machine to have the mechanically proper function of wrapping *gifts*?

I think not. Part of the reason is that the machine, *as designed*, cannot be a gift wrapper without being also a parcel wrapper or a unit packager. So all that remains to identify the "proper" function of the machine is the initial goal of the shopkeeper. Millikan (1984) thinks that this is enough. She believes that entirely new things can have functions *derived* from their creators' intentions. But I want to suggest that the shopkeeper's initial purpose has less of a claim to being the determiner of the function of the machine than the machine's physical constitution. Thus if someone were to insist that the machine's proper function is to wrap gifts because that is what the shopkeeper *had in mind* when he designed it, one cannot help but feel a certain arbitrariness in this claim. The whole idea of a mechanically proper function seems, in this case, to be an artificially imposed classification which dogmatically favors *one* of the tasks the machine can perform over others that are also possible — in fact, unavoidable — in virtue of a single underling architecture. The gift-wrapper might as well be a parcel-wrapper or a unit packager, for it can do all these jobs equally well. So while it is obvious that the machine "evolved" a design and that it has various capabilities stemming from this design, it is not so clear that it can be said to have a mechanically proper function, at least not if this function is to be decided by a mere inclination of the shopkeeper's will.

If we want to know what the machine can do without arbitrariness or prejudice, we must set aside the details of its genesis. We would then discover that the state of wrapping a gift is *functionally identical* to the state of wrapping a parcel and the state of forming a multi-item package, so that if the machine had a "functional point of view," as it were, it would not be able to differentiate these states: *it is doing just as much what it is supposed to do when it is wrapping gifts as when it is performing these other tasks.* That is not to say that the shop-

keeper's original plans are irrelevant absolutely, but it does suggest that they are tangential to the kind of work the machine is able to carry out *once designed.*

The same is true for human cognitive mechanisms. In fact, in the case of humans the point becomes more urgent, since we cannot even speak of the process of natural selection as having anything like the intentions of the shopkeeper. We can agree that people have the biological design they have because of certain selectional pressures in their evolutionary history, and that due to these pressures the human brain has acquired the capacity to map the world accurately. But the project of finding out what certain human brain structures do — insofar as such functions bear on the perspective of agents — is neither compromised by normative considerations nor restricted to causal–evolutionary facts. Dennett tells us exactly how to extract the functional properties of the brain and contemplate them in relative isolation: consider the brain as a novel artifact and find out what it can do from the way it is put together.

A Script for Narrow Content

I want to propose a way (similar to Dennett's in some respects but different in others) of capturing the narrowness of a thinker's intentional states — that dimension of an agent's beliefs, desires, fears, etc., which constitutes the agent's point of view. To that end, let us think of Adam's concept of water as a particular *script* (Schank and Abelson, 1977) — specifically, a *definitional* script (script_D). Despite what the name might suggest, scripts are not *linguistic* models, at least not *necessarily* linguistic. A script is best understood as a *schema.* This means that insofar as schemata can be non-propositional (and they surely can be), scripts can be so as well. In fact, scripts represent many different knowledge domains. There are scripts for personal stereotypes and functional roles, scripts for goal-oriented actions and common event sequences. There are scripts for spatial relations, personal habits, inanimate objects, living organisms, and even persons and selfhood. Moreover, scripts may — and often do — include visual and acoustic information, olfactory, gustatory, and other "purely phenomenal" data. In all cases, the content of a script is highly structured, and not simply a list of features or properties. Schank and Abelson (1977) confine most of their discussion to *episodic* scripts — those storing knowledge of generic events-types and workaday situations (e.g., going to a restaurant, visiting the doctor's office, taking the dog out for a walk). But it is natural and proper to extend their script notion to cover a variety of different conceptual categories. Thus a *definitional* script (an example of such an extension) is a mental schema that stores information about the identity and nature of *physical entities.*

I take it that scripts_D are fashioned in something like the picture Dretske (1981, 1988) paints for concept formation. Hence:

I. Definitional scripts are the product of analogue-to-digital data transformation. What is fashioned when one learns to recognize an object *o* as belonging to a certain type *T* is a script_D of *o*, and that script_D is to be understood as an internal network of neural circuits that have acquired, through prior experiences of the agent and local learning episodes, selective sensitivity to those properties of *o* that make *o* a token of *T*. Thus the process by which analogue information about *o* becomes digitally focused on the *T*-ness of *o* is the process by which certain neural structures adapt so as to react discriminately to specific features of *o*.

II. There will be a *loss* of information resulting from the digitalization process. This loss is essential to the extent that the emerging script_D acquires the job of classifying the input as a token of a generic type. Although classificatory organization greatly expands cognitive capacity, there is a trade-off. *A loss in informational content means a loss in the discriminatory power of the information-processing system*, and the effect is proportional. The more data lost, the more diluted a system's ability to discriminate becomes.

On the present proposal, then, scripts_D *are* the Dretskean neural configurations that arise in consequence of analogue-to-digital conversion. We may call the neural substructures in a definitional script — those responsible for isolating individual pieces of information in a stimulus type — the *frames* of that script_D, so that when enough of these frames become "excited" or "stimulated," object recognition (the recognition of *o* as *T*) is achieved.[8]

Now, what exactly is Adam's water-definitional-script ($\mathbf{WATER}_{DS}$) supposed to do? The causal–evolutionary reply is familiar: Adam's $\mathbf{WATER}_{DS}$ picks out water (H_2O) because Adam's history ties him to that particular liquid and to no other. But there is a better answer: the function of Adam's $\mathbf{WATER}_{DS}$ is to pick out exactly those *substances* which manifest all the features to which the frames of Adam's water-script_D are sensitive. After all, Adam's $\mathbf{WATER}_{DS}$ picks out H_2O *by picking out a certain set of salient characteristics of the liquid* and, therefore, picking out those characteristics is as much a fact of Adam's evolutionary development (and local learning episodes) as the selective pressures his ancestors were under.

[8]See Marr's (1971) theory of *autoassociation*.

The situation here is analogous to the gift-wrapping machine. The function of the machine is to wrap items falling within specific physical parameters. As long as these items fall within the specified parameters, the fact that the items may be gifts, or mail-order catalogues, or multi-unit packages is *extraneous* to the manner in which the machine performs its work. Similarly, the function of Adam's **WATER**$_{DS}$ is to pick out substances that manifest particular physical features. As long as these substances display the features in question, the ultimate nature of the substance picked out will have no bearing on the functional performance of the script$_{D}$. To be sure, there may be many distinct objects that share all the features to which Adam's **WATER**$_{DS}$ is sensitive, thus eliciting a singular, undifferentiating scriptal reaction. But in behaving uniformly to each of these objects, Adam's **WATER**$_{DS}$ will be carrying out exactly the job for which it evolved — the job of responding selectively to item(s) with peculiar properties or characteristics, not *necessarily* pointing an identifying finger at H_2O, XYZ, or any other relevantly similar compound. In this way, functional indeterminacy intrudes upon the question of content.[9]

So what we have, on the one hand, is that thing to which Adam's "water" experiences have historically been causally connected (i.e., H_2O) and, on the other, the *sort* of thing that Adam's **WATER**$_{DS}$ picks out or reacts to in consequence of the way it is put together. The scriptal approach I am advertising thus suggests a bifurcation of intentionality into at least two distinct dimensions. Intentional *aboutness* is a causal relation that connects an agent to some external object — either directly or indirectly — and may therefore be analyzed historically–evolutionarily. But there is more. What I want to call intentional *directedness* is a feature of thought that is dependent on nomic relations between properties of objects and properties of definitional scripts. Adam's belief is directed toward water because the property of *being water* is nomically connected to the frames of his water-script$_{D}$ in virtue of what these frames indicate. Furthermore, for *any* object *o* that answers to all the frames of Adam's script$_{D}$, the property of *being o* is nomically connected to Adam's **WATER**$_{DS}$. It follows that a nomic relation between an object and a definitional script can be in place even if that object has never previously been a cause of that script$_{D}$'s priming or activation. In short, teleology *can* be used to ground intentional aboutness, but *only* intentional aboutness. If it is Adam's perspective we are after, if it is his *narrow* intentional state we seek to illuminate, we shall have to take notice of more than just the things that

[9]There remains the issue of misrepresentation, which I do not have the space to discuss here. Suffice it to say that a rich account of misrepresentation — an account that does *not* rely on the evolutionary or causal history of the misrepresenter — can be given by comparing what I call (below) the "aboutness extension" and the "directedness extension" of an intentional state.

have historically elicited his thoughts; we shall have to take notice of the things to which these thoughts are directed.

Whereas aboutness is meant to capture the *referential origin* of an intentional state, directedness is meant to capture the *range of the agent's intentional point of view*. Another way to say more or less the same thing is that aboutness is fixed *widely* whereas directedness is fixed *narrowly*. Whether or not an intentional state is *about* an exclusive object depends on what connects the state to the world. But whether or not an intentional state is *directed* toward an exclusive object depends on what the agent, as an information-processing system, is capable of discriminating under a specific script_D. Aboutness and directedness can therefore give rise to different extensions: a state may be about x (have x as its *wide content*) but be directed toward y (have y as its *narrow content*), where $x \neq y$. In order to find out what an intentional state may be about and to what it may be directed, we shall construct for each state an *aboutness extension* (E_a) and a *directedness extension* (E_d).

What goes into the E_a of an intentional state is whatever real or actual object bearing relation C to the agent whose intentional state it is. C is a *causal* relation, in consequence of which a representation is primed. The primed representation is a particular script_D which is a product of a system's ability to digitalize incoming analogue information. The aboutness of an intentional state, then, may be defined in terms of a causal relation connecting a real-world object with a *primed* definitional script in an agent. The connection need not be direct, but the etiology of C must trace back to a real object in order for the relevant intentional state to have aboutness.

The absence of aboutness does not mean that a mental state is not intentional, however, since aboutness, as it is conceived here, is not a *necessary* component of intentionality. To say that aboutness is not necessary for intentionality is merely to acknowledge that one can truly believe that Santa Claus is fat and jolly, that one can have a genuine desire to meet the Tooth Fairy, that one can be really fearful of the Wolfman, even though none of these entities exists. And the intentional status of such beliefs, desires, and fears is not in any way compromised on account of their aboutlessness.

What goes into the E_d of an intentional state is *any* object that answers to the frames of the script_D underlying the state. We shall adopt a *modal* reading of "any" here — a reading that extends the meaning of the word to include possible but non-existent items. To do otherwise would be to exclude aboutless intentional states involving the kind of fictions or fabrications mentioned above. The modal reading thus allows us to have a unified treatment of the directedness of all intentional states, those pertaining to real objects and those that do not. What I mean by an object "answering to" the frames of a script_D is simply the frames' *reacting causally* to the object in a counter-

factual way: the neural structure constituting the script$_D$ *would* respond to the object *if* the object were real and present.

Notice that this account of directedness (unlike aboutness) is sensitive to what a system can and cannot *discriminate*. This is in keeping with our endeavor to better understand the point of view of the agent. Thus Dretske (1995) claims that whether or not two concepts are identical is "not simply a matter of knowing, or not knowing, the right labels or words for experienced differences. It is, instead, a matter of lacking [or having the relevant] discriminatory powers" (p. 138). Fodor (1994) makes a similar point when he comments that if "a creature can't distinguish Xs from Ys, it follows that the creature can't have a concept that applies to X but not to Y" (p. 32). We can articulate the same intuition as the *principle of intentional inclusion* (PII).

PII: An information-processing system (agent) A cannot be in a script-based intentional state I_s directed *exclusively* toward an actual or possible object o_i if A is incapable of discriminating between o_i and other superficially similar actual or possible objects, o_j, o_k, o_l, ...

The reasoning behind PII is that, since the object(s) which go into the E_d of an intentional state are determined by a script$_D$, that intentional state does not identify in those objects any differences that cannot be identified by the script$_D$ itself. In other words, where A is an information-processing system, I_s a script-based intentional state of A, and $E_d(I_s)$ the directedness extension of I_s, PII essentially limits the individuation of all $o \in E_d(I_s)$ to the range of discriminations made by A based on A's activated script$_D$.

What we now need — what is long overdue, I believe — is a characterization that renders the whole idea of narrow intentional content much more precise and robust than it has been traditionally conceived. Narrow content, I propose, just *is* the directedness extension of an intentional state, so a definitional statement of $E_d(I_s)$ is what we are after. Such a statement would provide us with the perspective of the thinker in tidy and lucid terms.

In order to say with precision what a directedness extension is, we shall use PII to construct an *equivalence class* of actual and possible objects relative to an arbitrary definitional script S. An equivalence class is determined by means of an *equivalence relation*. Mathematically, a relation $\mathbf{R}$ on a given domain Ω is an equivalence relation on Ω if and only if (iff) $\mathbf{R}$ is *reflexive* on Ω, $\mathbf{R}$ is *symmetric* on Ω, and $\mathbf{R}$ is *transitive* on Ω. $\mathbf{R}$ is reflexive on Ω iff for all $o_i \in \Omega$, $o_i \mathbf{R} o_i$. $\mathbf{R}$ is symmetric on Ω iff for all o_i and $o_j \in \Omega$, if $o_i \mathbf{R} o_j$, then $o_j \mathbf{R} o_i$. $\mathbf{R}$ is transitive on Ω iff for all o_i, o_j, and $o_k \in \Omega$, if $o_i \mathbf{R} o_j$ and $o_j \mathbf{R} o_k$, then $o_i \mathbf{R} o_k$. Let us first define $\mathbf{R}$ and Ω:

Let S be an arbitrary script$_D$ and $F = \{f_1, f_2, f_3, \ldots, f_n\}$ be the set of frames constituting S. For any two actual or possible objects o_i and o_j, we say that $o_i \mathbf{R} o_j$ iff for every $f_i \in F$,

both o_i and o_j possess the *feature* or *characteristic* or *property* to which f_i is selectively sensitive. Now we let Ω be the set of all actual or possible objects for which the relation **R** holds.

The first thing to point out is that Ω is a well-defined set. Well-definedness neither entails that a set be finite nor that its members be corporeal objects. Rather, what well-definedness requires is that there be a *decision procedure*, a definite "yes" or "no" answer to the question whether the item being contemplated belongs in the set. That is precisely what we have for Ω. Any actual or possible object o_i will be a member of Ω iff for every $f_i \in F$, o_i possesses the feature for which f_i has acquired the job of indicating. Otherwise, $o_i \notin \Omega$.

Secondly, it is not difficult to see that **R** is an equivalence relation on Ω. For every $o_i \in \Omega$, o_i has the same properties as itself (including, of course, those properties to which all $f_i \in F$ are sensitive); so "o_i **R** o_i" is true (reflexivity). For any two objects o_i and $o_j \in \Omega$, if o_i has the same properties picked out by all $f_i \in F$ as does o_j, then o_j also has those properties in common with o_i; so "$(o_i \mathbf{R}\ o_j \rightarrow (o_j \mathbf{R}\ o_i)$" is true (symmetry). For any three objects o_i, o_j, and $o_k \in \Omega$, if o_i has the relevant properties in common with o_j and o_j has the same relevant properties in common with o_k, then o_i also shares those properties with o_k; so "$[(o_i \mathbf{R}\ o_j) \& (o_j \mathbf{R}\ o_k)] \rightarrow (o_i \mathbf{R}\ o_k)$" is also true (transitivity).

We now define the *directedness* of an intentional state using **R** as follows:

Let I be any intentional state, S be the script$_D$ underlying I, $F = \{f_1, f_2, f_3, \ldots, f_n\}$ be the set of frames constituting S, and o_i be any actual or possible object such that, for every $f_i \in F$, o_i displays the feature for which f_i is selectively sensitive. The equivalence class of o_i as determined by **R** is fixed by the set $o_i/\mathbf{R} = \{o_j \in \Omega \mid o_i \mathbf{R}\ o_j\}$.[10]

In plain English, what $o_i/\mathbf{R}$ does is pick out, for any intentional state involving a definitional script, all actual and possible objects that possess the total sum of features or characteristics which the frames of the script$_D$ have the job of indicating. Hence, all of these objects will be *equivalent* or *identical* or *indistinguishable* from the perspective of the agent whose script$_D$ it is, because they are undifferentiated by the script$_D$ in question; which is to say that the world-according-to-the-agent is not one in which the members of $o_i/\mathbf{R}$ exist as separate entities. If we let "I_s" designate the appropriate script-based intentional state, then $E_d(I_s) = o_i/\mathbf{R} = \{o_j \in \Omega \mid o_i \mathbf{R}\ o_j\}$. Now we know exactly how to construct the set of possible worlds Dennett spoke of, the worlds where Adam cannot distinguish between *o* and *o*'s superficially similar cousins. *These will be the worlds that contain at least one member of* $o_i/\mathbf{R}$.

[10]This is read "the class of o_i modulo **R**" or simply "o_i mod **R**."

The way **R** is defined does not commit us to specifying the kind of features, or characteristics, or properties to which the frames of our arbitrary script_D may be sensitive. What **R** does is define an equivalence class *relative to the manner in which the frames of some* script_D *react*, regardless of what actual properties are being reacted-to by these frames. The external properties may be whatever you like; they may even be radically different from one object in the equivalence class to another. The important thing is that they must elicit the *same* response from the relevant definitional script, in which case the script_D would represent all of the corresponding objects as identical whether or not they are actually so. *That* is what it means to have an equivalence class *relative* to a definitional script. The equivalence of the members comes from the way in which their associated properties affect the script_D, not from the sort of properties they in fact have.

The claim proposed here is not that definitional scripts have the job of selecting some set of *features* or *properties*, but the job of selecting *objects* with particular features or properties. So, to return to our old example, it just so happens that the way Adam's $\textbf{WATER}_{DS}$ is physically realized in his brain makes it impossible for the script_D to pick out water (i.e., react to H_2O in a particular way) without also picking out water_{te} (i.e., react to XYZ in exactly the same way). The distinction between intentional aboutness and intentional directedness clarifies how Adam's "water"-thoughts may be linked to his environment via H_2O or XYZ, while simultaneously illuminating his perspective by defining the range of the discriminations he is able to make. This approach sidesteps concerns about functional indeterminacy by embracing it, by using scriptal imprecision to get a handle on how things appear from a thinker's point of view. But the problem of functional indeterminacy sill remains for causal–evolutionary theories: if Adam's $\textbf{WATER}_{DS}$ cannot distinguish H_2O from XYZ, on what objective grounds do we limit the *function* of the script_D to the identification of the one substance but not the other? Appealing to causal history only begs the functional indeterminacy question at issue.[11]

---

[11]It is worth mentioning that not all scripts_D manifest indeterminacy in the way they function. Suppose, for instance, that Adam encounters an object o_i which primes in him O_{DS} (where O_{DS} is Adam's definitional script of o_i). Thus whenever o_i primes O_{DS} in Adam, he will be in the intentional state I_s of recognizing (i.e., believe that he perceives) o_i. In this case, $E_a(I_s) = \{x \mid x = o_i\}$ since o_i stands in C to Adam; $E_d(I_s)$ will be the set $o_i/\mathbf{R} = \{o_j \in \Omega \mid o_i \mathbf{R} o_j\}$ — that is, the set of all actual or possible objects manifesting the properties to which the cluster of neuronal structures constituting O_{DS} has the job of indicating. Now suppose that Adam's definitional script of o_i is such that $E_d(I_s)$ contains just one element, o_i itself. Here we have a situation where Adam's o_i script_D is so precise that no actual or possible non-o_i answers to all the frames contained in it. When Adam has a belief about o_i, or a desire for o_i, or a fear of o_i, his intentional state will be directed toward exactly the same object that the state is about. Whatever else may be the case, then, we can rest assured that the totality of the frames of Adam's definitional script of o_i — if *all* these frames are implicated in his recognizing o_i — will not cause him to misidentify a token of a non-o_i for a token of an o_i. There is, in other words, no possibility of functional indeterminacy for O_{DS}.

Back to the Swamp

Davidson's (1998) "Swampman" thought-experiment illustrates the unacceptable cost of emphasizing aboutness (analyzed in terms of causal history) over directedness (analyzed in terms of dispositional nomic relations), thus revealing the second major challenge facing teleological accounts of intentionality. Imagine that lightening strikes in a swamp. Both Davidson — a respectable philosopher and a psychological externalist — and the tree under which he was standing are completely disintegrated. But while the original Davidson is lost, by a fantastic coincidence the molecules of the disintegrated tree are reconstituted into the physical replica of Davidson just as he was immediately prior to the accident. Swampman behaves exactly as Davidson does so that "no one can tell the difference" (Davidson, 1998, p. 91). But there *is* a difference. According to the causal–evolutionary view of intentionality, the difference is that Swampman has *no intentional states at all* — no beliefs, no desires, no hopes or fears. This is because he lacks, by assumption, the causal connections to the world on which the content of thought depends.

Fodor (1994) claims to have heard Millikan explicitly sanction this appraisal of the imagined Swampman. The verdict is also endorsed by Davidson himself and by Dretske, but unlike Davidson and Millikan, Dretske appears to accept it grudgingly. Dretske does not argue for it directly, but points out that a zombie-like Swampman is *possible* (1995, p. 148). We can acknowledge this without conceding very much. What is required to make the causal view credible is not only the bare *possibility* of a contentless-headed Swampman, but a *non-question-begging* argument telling us why we should take that possibility seriously. I believe we have absolutely *no* reason to think that Swampman is in any way different from the original Davidson.

Let us retain the same story except for one minor change. Suppose Davidson is not disintegrated by the lightening bolt, but is rendered unconscious for a short time. When he awakens, he notices another figure struggling to get to his feet. Each of the two Davidsons is shocked to see the other and does not quite know what to make of the situation, but they agree on some fundamental things. Each remembers a lightning bolt hit a tree just before he was rendered unconscious; each recalls seeing an "impostor" staring back at him when he regained consciousness; each claims of himself to be the "real" Donald Davidson and the other to be nothing more that a tree-product; holding a Davidsonian causal theory of intentionality, each insists that the other is a counterfeit who only *appears* to have genuine thoughts.

Davidson's family does not know how to tell the two apart, but they are naturally keen to find a resolution. Accordingly, they put together the finest scientific team they can manage — psychologists, sociologists, biologists, doctors and surgeons — together with some pertinent laypersons — friends,

acquaintances, former students, colleagues and kin. After some discussion, it is agreed that the best way to discover the "real" Davidson is to put the causal–evolutionary account of intentionality *on trial*, to *assume* that it is correct and then use it to figure out which of the two is the zombie.

Shortly thereafter, they start their investigation. Biologists take blood and tissue samples from the two to compare their DNA. Psychologists conduct lengthy interviews to try to uncover any anomaly that might throw suspicion on one or the other. Sociologists do their best to chart their relationships with their peers and the organizations to which they belong. Doctors and surgeons probe their bodies and compare the results with their records. Do they have this or that scar from a past surgery? Are their dental records the same? Do they suffer from the same ailments? Are they in need of the same medication(s)? Finally, their friends and family are allowed to question them. Do they remember such-and-such event, person, or circumstance? Does each seem like the Davidson they knew, with the same mannerisms, habits and sensibilities? If we accept the claim that "no one can tell the difference," the results of all these tests are *guaranteed* to be inconclusive.

At this point, we are entitled to wonder what possible grounds we have for accepting the claim that one of the doppelgängers does not really have intentional states. Davidson's good word is just not good enough! We have conducted all the tests we can think of and found absolutely no dissimilarity between the candidates. This negative result is not *neutral* with respect to the causal–evolutionary doctrine; it serves to undermine it. Do we not have good reason, *based on empirical evidence*, to reject a *necessary* link between the developmental origins of an organism and that organism's thoughts and attitudes?

We can look upon the procedures used to test Davidson and Swampman as a very elaborate kind of Turing test. What makes this Turing test elaborate is that we are not attempting to ascertain the presence of a mental life only by analyzing the way questions are answered, but by analyzing *everything* we can think to analyze: chemistry and biology, personal habits and preferences, social relations, moral character and religious inclinations, philosophical opinions, and whatever else you like. The trouble is that by premise, *nothing we could ever do* will reveal a disparity. In the face of this kind of futility, what sense can we even make of the claim that there *still is* a difference? How might this difference be brought to light, and who exactly has the burden of proof in this debate?

This problem has forced Fodor (1994), even in his externalistic mood, to admit that Swampman is a "serious embarrassment" for historical accounts of intentionality:

> Of course, not having had one, Swampman doesn't remember his twelfth birthday party; "remember" is factive, and you can't remember what didn't happen. But it seems

> very odd to say that Swampman doesn't know what time or day of the week it is . . . . If it's not his believing that it's Wednesday that explains why the Swampman says "It's Wednesday" when you ask him, what on earth does? . . . To put the point another way: Perhaps it's true, as it were, by definition that beliefs, desires, lusts and the like are constituted by their histories; in which case, of course, Swampman doesn't have them. But, so what? It's intuitively plausible that he has states that are their exact ahistorical counterparts *and that these states are intentional.* (p. 117)

I am inclined to go further. I think Swampman amounts to a *reductio ad absurdum* of this family of theories. The brand of externalism which derives from a tendency to regard evolutionary history as essential for content is, I think, misleading at best. And yet, there is nothing in what I have said about intentionality that is inconsistent with the idea that our history and environmental conditions exert great influence on our mental lives. In fact, the entire notion of scripts is based on a presumption of a complex network of interactions between agents and world. So far, anti-individualist intuitions are well accommodated.

On the other hand, it is important to understand that the environment exerts its influence by proxy, through the representational medium of the physical brain. Intentional directedness is produced by networks of neural structures that are sensitive to specific types of stimuli. What gives these structures intentional content is *that* they are thus sensitive, not *how* they became thus sensitive. Perhaps that is what Fodor had in mind when he suggested that Swampman could have *a*historical intentional states. At any rate, Swampman has the same neuronal structures as Davidson, and insofar as these structures constitute scripts in one individual, they constitute scripts in the other. How these scripts came about is not nearly as important as the fact that they are there.

But because we *are* products of a long and complex evolutionary history, we should resist radical internalism. To really appreciate the intentionality of thought, we must come to terms with both its aboutness *and* its directedness. To that end, we can think of an intentional state I_s in terms of the ordered pair $\langle \mathbf{A}, \mathbf{D} \rangle$, where $\mathbf{A} = E_a(I_s)$ and $\mathbf{D} = E_d(I_s)$. Doing so permits us to concentrate on one or another intentional dimension as best suits our exploratory ambitions: we can consider $\mathbf{A}$ independently of $\mathbf{D}$, and $\mathbf{D}$ independently of $\mathbf{A}$. But if our goal is to achieve an understanding of intentionality pure and simple, $\langle \mathbf{A}, \mathbf{D} \rangle$ must be contemplated as a single unit, and in so doing we shall get as close to the exact character of I_s as we can ever hope to get.

References

Burge, T. (1998). Individualism and the mental. In P. Ludlow and N. Martin (Eds.), *Externalism and self-knowledge* (pp. 21–83). Stanford, California: CSLI Publications.

Crane, T. (2003). The intentional structure of consciousness. In Q. Smith and A. Jokic (Eds.), *Consciousness: New philosophical perspectives* (pp. 33–56). New York: Oxford University Press.

Davidson, D. (1998). Knowing one's own mind. In P. Ludlow and N. Martin (Eds.), *Externalism and self-knowledge* (pp. 87–110). Stanford, California: CSLI Publications.

Dennett, D. (1987). Beyond belief. In D. Dennett, *The intentional stance* (pp. 117–202). Cambridge, Massachusetts: MIT Press.

Dretske, F. (1981). *Knowledge and the flow of information*. Stanford, California: CSLI Publications.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, Massachusetts: MIT Press.

Dretske, F. (1990). Misrepresentation. In W. Lycan (Ed.), *Mind and cognition: A reader* (pp. 129–143). New York: Blackwell.

Dretske, F. (1995). *Naturalizing the mind*. Cambridge, Massachusetts: MIT Press.

Fodor, J. (1994). *The elm and the expert: Mentalese and its semantics*. Cambridge, Massachusetts: MIT Press.

Kripke, S. (1980). *Naming and necessity*. Cambridge, Massachusetts: Harvard University Press.

Marr, D. (1971). Simple memory: A theory of archicortex. *The philosophical transactions of the Royal Society of London (Series B, Biological Sciences)*, 262(841), 23–81.

Millikan, R. (1984). *Language, thought, and other biological categories: New foundations for realism*. Cambridge, Massachusetts: MIT Press.

Millikan, R. (1993). Thoughts without laws. In R. Millikan, *White queen psychology and other essays for Alice* (pp. 51–82). Cambridge, Massachusetts: MIT Press.

Putnam, H. (1975). The meaning of "meaning". *Minnesota Studies in the Philosophy of Science, 7*, 31–93.

Schank, R., and Abelson R. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, New Jersey: Erlbaum.

Segal, G. (2000). *A slim book about narrow content*. Cambridge, Massachusetts: MIT Press.

Stalnaker, R. (1999a). Twin Earth revisited. In R. Stalnaker, *Context and content* (pp. 210–21). New York: Oxford University Press.

Stalnaker, R. (1999b). On what's in the head. In R. Stalnaker, *Context and content* (pp. 169–193). New York: Oxford University Press.