

## **Function: Under Construction**

Eric Lormand

University of Michigan

August, 2008

Philosophers face many vexing moral problems. Though I insist several aren't my fault, maybe this paper'd better not go there. Sometimes these problems depend on puzzles about conscious experience, such as which beings feel pain, and when. This paper is barely sensitive to these puzzles. Consciousness seems to depend on mental representation or meaning, on what it is for a mental state to be "about" some topic. Here at last we have arrived at what this paper is also not really about. What something means seems to depend on its teleological "functions"—what it is "supposed to" do, what it is somehow "defective" if it doesn't do. And understanding functions is Job One here. I'll describe a general "constructive" theory of functions which I think improves and unifies the dominant heretofore competing theories. We'll get to Job One second, after raising the stakes via consciousness and its connections to morals and meaning.

### ***Job Two***

"How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of Djin when Aladdin rubbed his lamp." That was the pessimistic theoretical assessment at the dawn of scientific psychology, from the otherwise famously optimistic biologist Thomas Huxley, known as

“Darwin’s Bulldog”. Despite the many subsequent scientific advances in understanding the mechanisms of perception, thought, and communication, and despite the many subsequent philosophical advances in understanding the nature of mental representation and meaning, conscious experience is still The Rub.

Contemporary Doubting Thomases converge on a general characterization of the mystery. Thomas Nagel writes that “[c]onsciousness is what makes the mind-body problem really intractable”, identifying its most troublesome feature as follows:

[F]undamentally an organism has conscious mental states if and only if there is something it is like to *be* that organism—something it is like *for* the organism. We may call this the subjective character of experience ....

Joe Levine similarly insists that “what is at issue is the ability to explain qualitative character itself; why it is like what it is like to see red or feel pain”. It’d be good to give clear examples of conscious states beyond seeing red and feeling pain, to contrast these with clear cases of nonconscious states, and to give some layout of the controversial cases. Unfortunately there are precisely zero noncontroversial candidates for nonconscious states, and a surprisingly long list of controversial candidates for conscious states.

### *Unclear consciousness and morals*

Here’s the quickest and most sweeping way to see that there are no untendentious examples of nonconscious states. Since it is so difficult to explain experience as a combination of nonexperiential things, some influential philosophers have been driven, by respectable argument,

to the view that experience is a *fundamental* feature of the world, like gravitation or electromagnetism or whatever underlies them. But just as virtually everything exerts fundamental forces like gravitation and electromagnetism, so if experience is fundamental we might well expect virtually everything to have experiences—even cells and rocks and raindrops and atoms. This is called “panpsychism” or, more precisely, “panexperientialism”. To be fair we can’t conveniently rule panpsychism out using bedrock assumptions about which entities, if any, lack experiences.

To zoom in on states that are uncontroversially conscious experiences, we have to zoom in on creatures that are uncontroversially conscious *experiencers*. This calls for postponing discussion of:

1a: Nonbiological beings such as laptops, robots, deities, and ghosts—(potential) robots *seem* nonconscious because they aren’t alive, but (potential) spirits *seem* conscious despite not being alive.

1b: Nonhuman biological beings such as plants, animals, and extraterrestrial life—some animals *seem* to be experiencers (down to bugs? shrimp? lizards? pigeons? bats?), but perhaps a degree of self-consciousness is required for full-fledged experience, and perhaps only language-users qualify. At the other extreme, vegetarians tend to fall silent when roast-beast eaters ask how they know plants don’t feel pain (except to call attention to the additional plants it takes to feed the animals we eat).

1c: Human beings far from adulthood, such as fetuses, infants, and small children—some of the dispute about abortion turns on whether fetuses suffer, which is an intuitively open question. It would be a symptom of pathology intuitively to deny experience in the face

of a normal newborn or a toddler, but to deny this based on the theoretical need to explain self-consciousness or “childhood amnesia” is merely a symptom of philosophy.

1d: Human beings of age but under extreme medical duress, such as comas, brain trauma, or anesthetic drugs—we entertain tales of near-death *experiences* from the nearly brain-dead, and we worry (with indirect evidence that this has often happened) that anesthetics merely paralyze us and then make us forget the excruciating experiences.

1e: Medically healthy human beings whose minds have undergone extremely limited developmental paths, such as people extremely mentally impaired or people raised by wolves—there are live theories according to which consciousness is a sophisticated cultural achievement rather than a primitive biological achievement.

1f: Apparently nonactual beings culled from philosophical thought experiments, such as alleged zombies, group minds, and homunculi or allegedly sentient subsystems within minds--as usual, theories and theorists disagree squarely about the proper assessment of such thought experiments, about what it is intuitive to say about imagined cases, and about whether the cases play tricks on intuitions.

That leaves, as noncontroversial experiencers, only Human, Old-enough, Minimally-healthy, Incarnate Experiencers, or “Homies”, for short. I’m a Homie and I write as if you are too. If you doubt there are Homies besides yourself, feel free to treat the plural forms of “Homie” and “experiencer” as typos, and feel free to treat “we” and “us” as applying specifically to Your Highness.

Now, what sorts of Homie states are uncontroversial examples of experiences? Here I also count six categories of nonstarters, states too theoretically charged for us to begin with:

- 2a: Intuitively nonmental states, including both “outer” states such as one’s location and hair color, and “internal” states that contribute to behavior, such as the conditions of one’s digestive system, one’s atomic particles, and one’s brain chemicals—panpsychists can treat any of these states as experiences, and physicalists treat many brain states as mental states in disguise, and so as potentially experiential.
- 2b: Aspects of our minds that are merely dispositional, such as forgetfulness, cleverness, and skills—since dispositions need not even be exercised to be real, it’s odd to imagine them constantly participating in our swarms of conscious experiences, but perhaps they participate in a way that is hard to separate out (forming part of a constant background of experience, say), or perhaps the subsystems underlying these dispositions within our minds have these states as their own experiences, separate from “ours” (see category 1f of controversial experiencers above).
- 2c: Sensory-deprivation states such as those during sleep (including dreams and sleepwalking and sleeptalking), hypnosis, or deep meditation—perhaps dreams do not happen consciously during sleep, but instead dream scripts are nonexperientially smuggled into memory, so that upon waking we misremember ourselves as having had conscious dream experiences during sleep.
- 2d: Individual states that persist through sleep, even if they also manifest themselves during wakefulness, such as beliefs (or, interchangeably, judgments), desires, and other “propositional attitudes”—if an attitude is nonexperiential during sleep, plausibly (though controversially) it is also nonexperiential during wakefulness, perhaps merely being accompanied during wakefulness by separable experiences (e.g, visual-imagery states)

that do not individually persist during sleep. So wakeful attitudes inherit controversial status from their sleepytime forms.

I would also include emotions such as fear and hope in the category of attitudes that persist during sleep, and moods such as depression and happiness in the category of mere dispositions (to activate or suppress various emotions and other states). So emotions and moods are controversial cases to be postponed. Certain emotions and moods, and perhaps other states, might seem to be *hybrids* of dispositions, attitudes and experiential “feelings”, but in that case I assume it’s the “feeling” components that we need to focus on first, and not the hybrid states.

2e: States that are deeply hidden from introspection, such as Freud’s repressed unconscious, “blindsight” states that enable blind people to guess correctly about visual stimuli, retinal states and other “drafts” in the early stages of vision, Heidegger’s alleged primordial ways of being in the world, and Chomsky’s alleged grammatical know-how—someone who maintains that these are experiential can easily explain away contrary intuitions by the fact that the states aren’t introspectible.

2f: Introspectible states that aren’t clearly introspected, such as the fuzzy boundaries at the edges of our visual fields, the faint pressures around our bodies a moment before we attend to them, and fleeting, subliminal perceptions—maybe they contribute to experience; maybe they don’t; maybe they (or states in any of the other categories) are vague borderline cases that are not quite experiences and not quite nonexperiences.

Fortunately, there are uncontroversial *positive* examples of conscious experiences, so I think it’s best to search first for principles and theories that work for these clear cases. In dealing with the unclear cases (someday), I think it’s best to rely on the resulting theories rather than our shaky intuitions.

*Clear consciousness and meaning*

I group the short remaining list of clear conscious experiences into four overlapping categories, with the general provisos that these states are only clearly conscious when they are clearly introspected by Homies:

- 3a: (Mis)perceptions caused by sense organs, such as normal tastings or seeings of environmental objects, and degraded appearances of afterimages or ringing-in-the-ears. (Mis)perceptions differ from beliefs in being tightly bound to stimuli and behavior (such as guiding visuomotor skills), and being limited in reasoning and in subject matter.
- 3b: Bodily sensations, such as diffuse sensations of warmth or muscular fatigue, and more pointlike pains, tickles, or itches.
- 3c: Perception-like imaginings, such as voluntary envisioning or “replaying” of one’s own actions or perceptions, and involuntary nonlucid hallucinations—but excluding dreams, due to their occurrence during sleep. (The controversial status of dreams does not infect waking imaginings, because individual waking imaginings do not persist as dreams.)
- 3d: Elements of “streams” or “trains” of thought, such as verbal talking to oneself, or reading with the mind’s eye, and nonverbal thinking in pictures or making predictions by running through a (scale) model in one’s head. Such thoughts differ from beliefs—for example, although a single belief may persist while one is sound asleep, no single thought does.

What does consciousness require? The first sign that there may be some dependence of consciousness on meaning is that all these clearest cases of conscious experience are clearly

representations. (It's less plausible that meaning depends on consciousness, because not all clear mental representations are clearly conscious.) Perceptions and misperceptions typically indicate allegedly objective features of external objects. For example, a visual sensation might represent a certain spectral-reflectance feature or a shape feature at a certain place, or might categorize an object on the basis of many such features. Experiences of pressure, warmth, or limb-position involve bodily (mis)perceptions of pressure, warmth, or limb-position, while in pain experiences we (mis)represent parts of our bodies as throbbing, burning, stabbed, pounded, pinched, pulled, etc., and in each tickle or itch experience one represents parts of one's body as being rubbed or pricked with very specific intensities, directions, speeds, and contact-point sizes. (Orgasms are left as an exercise for the old-enough.) Imaginings plausibly involve the same representational mechanisms involved in the later stages of perception and bodily sensation (i.e., the stages past sense organs and limited initial processing). And streams of thought seem to be constituted by imaginings of words or of speech acts, typically auditory or visual ones, as well as imaginings representing nonverbal items. It seems difficult to imagine a conscious state that represents nothing at all. And it seems difficult to vary what an experience is like without varying the experience's subject matter, or how it represents its subject matter. (Some philosophers have tried, but their thought experiments take us far from clear cases.) So it's tempting to view the representational features of experience as prerequisites for the conscious features of experience.

A second sign that there is some dependence of consciousness on meaning is that in all the clearest cases, conscious experiences are themselves represented by their bearers. A clearly conscious experience not only makes its subject matter seem to exist (with certain features), but also itself seems to exist (with certain features)—seeing blurrily or clearly, hurting severely or



mildly, forming images in one's head or body, etc. As Nagel emphasizes (above), there is not only something it is like to see red or feel pain—a feature of the seeing or the pain—but something it is like *for* the organism. Since the state itself is obviously in the creature's *possession*, “for” would be redundant unless it means in the creature's *view*, so the requirement is plausibly that there is something the state is like *as represented* by the creature. Theorists differ about whether a creature introspects the state by thinking about it, or more primitively by inwardly perceiving it, or by having the state somehow be in part about itself. Theorists also differ on whether consciousness requires ongoing introspection or mere ease of introspectibility. But in any case for a creature to have clearly conscious experiences it must be capable of mental representation.

### *Meaning and function*

Now, what does mental representation require? The first sign that there is some dependence of meaning on teleological function is the nature of *misrepresentation*, of a misfit between what is meant and what is there. True beliefs, accurate perceptions, and fulfilled desires seem to meet some norms of *success* that are *failed* by false beliefs, misperceptions, and unfulfilled desires. A belief that it's raining *ought* to go away when the rain does. These norms aren't statistical norms (as if trillions of falsehoods were better than one), and aren't moral norms (as if commandment XI were “thou shalt not miss thy guesses”). And it would be difficult to explain them as practical norms about means to one's ends, since any desire for an end is itself a success or failure in the sense to be explained. So it's tempting to try explaining the relevant norms in some fourth sense, as functions. Functions also aren't statistical patterns (as if most

eggs and sperm developed into offspring), and aren't moral principles (as if warped straightedges were crooked). And it would be difficult to explain functions as practical means to desired ends, since any desire for an end seems itself to have teleological functions in the sense to be explained.

On the other hand, some mental representations misrepresent without any failure or malfunction. Consider—and remember for later—the philosopher's Swampman, a particle-for-particle match for a Homie, suddenly and accidentally formed when lightning struck the muck. It would be a malfunction to *believe* in Swampman, since there's no such being, but it isn't a failure to have an *idea* of Swampman, or a defect to *imagine* that swamp lightning formed your physical twin. Nevertheless such representations have their own semantically-relevant teleological functions: under various circumstances ideas are supposed to help form beliefs and desires with related meanings, and under various circumstances imaginings are supposed to help test or practice the formation of beliefs and desires with related meanings. So it remains plausible that for a creature to have mental representations it must be capable of having teleological functions.

A second sign that meaning depends on functions is that both share many eerie features regarding how they enter into apparent explanations. (It's less plausible that function depends on meaning, because not all things with functions have meaning.) Since this moves us squarely into discussing functions, I describe it in the next section.

## *Job One*

To emphasize what seems eerie about teleological explanations (and meaning-explanations), I'll contrast them with a pair of important, general, but familiar explanatory relations:

4a: Constitution—For example, Angell Hall has as spatial “parts” giant stone Legolike blocks, has as temporal “slices” various Angell-Halls-on-a-Graduation-Day, and has as “aspects” greyness and quarantining philosophers. All these are kinds of constituents.

4b: Causation—For example, laying the Legos increased the pressure on the pillars, and blocked the upward mobility of the quarantined.

I'll use “construction” to cover both constitution and causation. Throughout, I mean these broadly to include constituents and causes that work jointly with others, some perhaps contributing extremely little to the whole or to the effect. So “construct” stands for “help somewhat to constitute or cause”.

Now for the eerie contrasts (shared by teleological and meaning explanations):

5a: Backward explanation--Barring eeriness within the constructive realm, later things can't construct (constitute or cause) earlier things. But we routinely try to explain the existence of a belief or desire by properties of a future thing believed or desired, and the existence of an organ or tool by the future fulfillment of a function.

5b: Circular explanation—Barring eeriness within the constructive realm, things can't be constructed by things they help construct. But we routinely try to explain the existence of a belief, desire, organ, or tool by properties of a thing believed, desired, or had as a

function, even while we try to explain the existence of that thing as an effect of the belief, desire, organ, or tool.

5c: Absent explanation—Things can't construct what doesn't exist (at some time), and things can't be constructed by what doesn't exist (at some time). But we routinely try to explain the existence of a belief, desire, organ, or tool by alleged properties of the thing believed, desired, or had as a function, even when that thing never comes to exist.

5d: Explanation at a distance—Except in the most fundamental cases involving the tiniest things and spaces, we expect there to be deeper and deeper mechanisms explaining how things construct other things. But we routinely cite evidence and principles of reasoning as explaining “why” we perceive and believe what we do, and cite desires and intentions as explaining “why” we do what we do, without our knowing or much caring whether these lead to or fit with deeper explanations of “how”. And we routinely cite functions as explaining why objects do what they do, with little knowledge or care about whether these explanations lead to or fit with explanations of how they do what they do.

### *Function and history*

So what does teleological function require? It is obscure how we can talk about functions in a scientifically respectable way. If one thinks of the universe as so many randomly or lawfully crashing particles, it's hard to resist the idea that the universe just is the way it is, and there are no “shoulds” or “functions” there for the scientist to describe. Nevertheless there seems to be a sharp break among scientists who act like they don't need teleological notions, and scientists who act like they do need them. Using broad categories, that break falls between

microphysicists, chemists and macrophysicists on the “don’t” side, and biologists, social scientists, and engineers on the “do” side. (We can speak of atoms or molecules or volcanoes “functioning” in various ways, but this is a pun that means little more than that they have effects. It isn’t associated with the key *teleological* talk of them as succeeding or as failing, malfunctioning or being defective.) Functions seem to enter the universe when reproduction and selection does—with living beings and their artifacts—and this is the first sign that there is some dependence of function on evolutionary design.

“Backward-looking” or “etiological” theories of function accordingly try to show how histories of reproduction and selection generate teleological functions. Very roughly, the core etiological proposal is that some entity—call it *e*—has a certain function—call it *f*—if *e* exists because it is “copied” from similar ancestors that actually did *f* and thereby increased their probabilities of reproduction in competition with other things. Etiological theories treat functions as objective, at root determined independently of the theorist’s interests or of any other mental representations, and so make functions available for a reductive explanation of mental representation. And in Ruth Millikan’s hands, most impressively, etiology is carefully construed to unify biological and nonbiological functions, handling the functions of genes, body parts, instinctive behaviors, tools and other artifacts, customs, linguistic devices, mental states, and so on.

Etiological theories dispel the eeriness of teleological explanation by recasting it as constructive explanation. It’s not that some future or inexistent activity of an object explains, in a backward or absent way, the existence of an object. It’s that *past* occurrences of *similar* activities

explain—causally explain—the existence of the object. This introduces no circularity when the object causally explains the future fulfillment of its function. And instead of looking for gappy “why”-explanations distinct from “how”-explanations, we can look for 2<sup>nd</sup>-order how-explanations: not “how (now) does this mechanism work”, but “how (in the past) did such a mechanism come to exist (or proliferate)”? That causal question can sometimes be answered with little knowledge or care about how the mechanism works.

Since Swampman (see previous section) would have no ancestors, no history of selection, on the etiological theory it would harbor no teleological functions. This functionlessness is little surprise. Swampman’s pulsating heartlike tissue would not be defective if it stopped pulsating, since nothing relevant to it would have set a standard for it to fail. Its lunglike tissue doesn’t inflate by (artificial or natural) design, but as a pure accident of the lightning striking the swamp. Grinding food is not a *raison-d’etre* of its teeth-alikes. These structures are like interesting but functionless volcanic formations.

### *Job Two interrupts*

But one scary growl from the brush, and Swampman zips toward the distant city lights, finally reaches your house door—locked—then quickly enters your shed by punching in your secret combination code. Apparently he hears and fears the growler, wants to protect himself, believes safety lies toward the light, and knows the right shed combination. But since Swampman has no evolutionary history, on the etiological theory he has no states with teleological functions, so he has no meaningful mental states. If he has no state that malfunctions

if there was no growl, he does not have a perception (or misperception) of a growl. If he has no state that fails if he doesn't hide, then he has no desire to hide. He opens your lock by the sheer accident of being your molecular duplicate, not by knowing or even guessing what the combination is. This meaninglessness is all quite a surprise, but we can start to understand it. Just as there can be things that are *similar* to English sentences but that *aren't* really meaningful sentences (e.g., gibberish, or shapes and sounds formed by the wind, or ink merely spilled on a page—mere meaningless sentence-alikes), so there can be things similar to thoughts but that aren't really meaningful *thoughts* (e.g., perhaps data and programs in a laptop computer, or shapes and sounds and electrical discharge merely leaking in a soul or brain or mind—mere meaningless thought-alikes).

If it turned out we are *all* Swampfolks rather than evolved beings, there would be no thoughts at all, but only thought-alikes. A strange possibility, but one we can begin to get comfortable with, using the analogy of meaningless sentence-alikes. It wouldn't surprise us to discover this; it would merely surprise-alike us to discover-alike this. Of course no one would bother even trying to talk or think if nothing means anything, but only because no one would bother trying anything (versus bother-alike-ing try-alike-ing). The meaningless talk-alikes and thought-alikes would just continue to happen.

Now if consciousness in turn depends on meaning, consciousness also depends on a history of reproduction and selection. Swampman would have no conscious experiences; there would be nothing any of his states is like for him. This nonconsciousness is a huge surprise. Meaningless sentence-alikes look or sound intrinsically like sentences with functions; all that's

missing is the relational meaning. So it's tempting to treat Swampman's meaningless perception-alikes and thought-alikes as sharing Nagel's "subjective character" or Levine's "qualitative character" with conscious perceptions and thoughts; all that's missing is the nonintrinsic meaning about the external world. But without meaning and function there is also no *internal* world "for" a "subject". No perceiving eyes, and no introspecting mind's eye. We could describe some of Swampman's states as experience-alikes, but since there would be nothing it is like for him to be in these states, in conscious respects they would be no more similar to conscious experiences than our own nonconscious states are. And to the extent that certain moral rules stem from what we may not do (or must do, or should do) to those who feel pain, since there is nothing it is like for Swampman to feel pains (or to feel-alike pain-alikes), these rules would not apply to our treatment of him. Apparently it's not who your parents are, it's whether you have parents.

When we try to get our heads around the possibility that we are Swampfolks, it seems even more mysterious how our states could be wholly nonconscious than it is how our states could be wholly meaningless. Our continued assurance that we have *meaningful* states, in the face of the possibility of being Swampfolks, might be based on a subtle working assumption—we might as well act as if we have beliefs, since if we don't, at least we're not mistaken. Or it might be based on more elaborate reasoning, such as our considerable evidence that we are in fact evolved creatures. But our assurance that we have *conscious* states seems more directly supported than that. It's tempting to say we know we are conscious just because we are conscious. This knowledge seems independent of the environment outside us because the conscious experiences seem independent of the environment outside us. Our pains, and our



resulting moral status, seem to be constituted only by how our mind-brains are here and now. Or perhaps they are partly constituted by a little bit of the environment—aspects of our bodies, or of the past few seconds of stimuli and behavior. But—slap to the head!—how could whether *this* hurts depend constitutively on distant facts about my evolutionary history?

*Job One continues*

To secure functions for Swampman, and sometimes for other motivations, philosophers have tried three chief alternatives to the etiological theory, attempts to understand functions without dependence on evolutionary histories of reproduction and selection.

*Function without history*

First, John Bigelow and Robert Pargetter argue that functions stem from “forward-looking” survival-enhancing *dispositions*, rather than “backward-looking” survival-enhancing *prior activities*:

The etiological theory describes a character [= feature] *now* as serving a function, when it *did* confer propensities that improved the chances of survival. We suggest that it is appropriate, in such a case, to say that the character *has been serving that function all along*. Even before it had contributed ... to survival, it had conferred a survival enhancing propensity on the creature. And to confer such a propensity, we suggest, is what constitutes a function. Something has a (biological) function just when it confers a survival-enhancing propensity on a creature that possesses it.

So on this propensity theory Swampman's structures, states, and processes have functions *from the first moment* if they are disposed to increase Swampman's chances of survival.

Second, Robert Cummins dispenses with the teleological idea that functions stem from explaining existence or survival of function-bearers, and substitutes the idea that functions stem from contributions relatively simple things make to the relatively complex capacities of containing systems:

*x* functions as a  $\varphi$  in *s* (or: the *function* of *x* in *s* is to  $\varphi$ ) ... just in case *s*'s capacity to  $\psi$  [is appropriately and adequately accounted for] by, in part, ... the capacity of *x* to  $\varphi$  in *s*.

An account of capacities "becomes less and less appropriate, and talk of functions makes less and less sense" as "the relative complexity of the organization of component parts/processes that is attributed to the system ... becomes less and less significant". On this organizational theory, from the first moment, Swampman's complex capacities (to run, to open combination locks, etc.) can be accounted for by much simpler contributions of many of Swampman's parts, so these accounts describe the functions of these parts (or: how they function) in Swampman's overall organization.

Third, John Searle reverses the dependence relations described in Job Two, maintaining that functions depend on conscious meaning. He says “functions are never intrinsic to the physics of any phenomenon but are assigned from the outside by conscious observers and users”:

Thus given that we already accept that for organisms there is a value in survival and reproduction, and that for a species there is a value in continued existence, we can *discover* that the function of the heart is to pump blood .... If we thought the most important value in the world was to glorify God by making thumping noises, then the function of the heart would be to make a thumping noise ....

On this attributional theory, from the first moment, Swampman harbors whatever functions our desires require him (or his parts) to achieve. We aren't essential—Swampman's own desires, or those of a bystanding deity, could provide the requirements. (And Searle thinks Swampman could have desires independently of functional attributions, because meaning and consciousness don't after all depend on functions.)

Each of these three alternatives has some merit alongside that of etiology, and perhaps the dominant tendency is to a pluralism accepting versions of multiple theories as separately sufficient for functions. Perhaps “function” is ambiguous, and each theory is correct about a distinct concept of function. Perhaps different theories are best restricted to different sorts of function-bearers (say, etiological or propensity theories for biological organs, and organizational or attributional theories for artifacts). Lopping together an etiological theory with an independent nonetiological theory would enable Swampman to have functions (and meaningful representations, and consciousness), but doing so too quickly would be cheating ourselves. It

would give us no account of why we treat these different kinds of functions alike. It would give us no guidance about what other sources of functions there may be—if there are these four kinds of functions, why not also functions as activities we *dislike*, functions as propensities whose shortest representation in German *rhymes* with the shortest English representation of contributions to containing capacities, etc.? Pluralism without unity is for quitters.

### *Function and construction*

So I would first like to show how a general account of functions can generate improved versions of the main alternatives to the etiological account, and then use it also to derive the etiological account, unifying them all in the process. What I think functions have in common is their role in teleological explanation—the apparently eerie, apparently nonconstructive, backward/circular/absent/distant explanation described at the beginning of Job One.

Suppose  $e$  is an entity (object, event, fact, property, anything) with function  $f$ . The teleologist asks: why does  $e$  exist (at all, or with certain features)? The teleologist answers: because  $e$  does  $f$ . The “does  $f$ ” is clearly constructive—meaning “helps constitute or cause  $f$ ”. But since things malfunction, and teleology allows for absent explanation, “does  $f$ ” has to be construed softly, perhaps as “possibly or conditionally or with increased probability helps construct  $f$ ”. Since that’s a mouthful, I’ll risk misunderstanding and rely on the phrase “helps construct” to convey the softness. (Compare: we have no ready understanding of “I built the bridge, but in the end it didn’t get built” but we can easily understand “I helped build the bridge,

but in the end it didn't get built".) So the teleologist's explanation is that  $e$  exists because  $e$  helps construct  $f$ .

My first suggestion is that this "because" also stands for constructive explanation, rather than for some eerie *sui generis* explanation. The teleologist cites a fact involving  $e$  (a possibility, or conditional fact, or probability) as constructing  $e$  (perhaps together with other causes or constituents of  $e$ —I will omit this for brevity). So this fact involving  $e$  would have to exist, and have powers of causation or constitution, even before  $e$  comes to exist. We can understand this preexisting fact as being a fact about what *would* (possibly or conditionally or probably) happen *if  $e$  were to exist*. It will be useful to have a mnemonic and a special notation for this allegedly preexisting fact involving  $e$ . I'll call it the "funky fact" and write it with italics in funky braces. The teleologist accepts the formula:

*{if  $e$  were to exist,  $e$  would help construct  $f$ }* constructs  $e$ .

To avoid the counterfactual mouthful, let me simplify the expression to:

*{if  $e$  exists,  $e$  helps construct  $f$ }* constructs  $e$ .

My second suggestion is that at the most general level functions are whatever  $f$ s satisfy this formula. Hence the

**BOLD CONSTRUCTIVE THEORY OF FUNCTIONS:**

$e$  has  $f$  as a function if and only if: *{if  $e$  exists,  $e$  helps construct  $f$ }* constructs  $e$ .

The interesting thing is to understand the Ways such a preexisting funky fact involving  $e$  can construct  $e$ . I count five Ways for the construction to go, each of which relates closely to a

preexisting theory of functions, citing a factor that would, if in place, plausibly be sufficient for functions. If there are further Ways, I conjecture they will also cite arguable if not already familiar sources of functions. Or if not, they will exclude themselves by naturally suggesting further conditions in the same spirit, for a more cautious constructive theory of functions.

*The Way of Abstract Causation (Classical Teleology)*

Nowadays it is normal to deny that abstract entities (e.g., properties themselves, as opposed to concrete instantiations of properties) have causal powers over what happens in the concrete world. Socrates sounds crazy when he insists that “as a consequence of the good”—construed somewhat misleadingly as the property of being good—“existence and being are in things”. Today John Leslie sounds crazy—as crazy as anyone else who grapples with the problem of explaining why there is something rather than nothing—when he appeals similarly to causation by abstract goodness:

An absence of all [concrete] existents truly would be in a way tragic because there might have been a good situation instead. And a situation sufficiently good—one which wouldn't be rather a pity because something far better could be there in its place—could perhaps have an existence required not just ethically, but with creative effect. Without contradiction or other absurdity, perhaps, the goodness of something might bear responsibility for that something's being more than merely possible.

I agree there is room to argue (as Leslie does) for abstract causation; there is nothing absurd about it, and I can think of no systematic attempts to defend the bias against it. So imagine there are circumstances in which something's being good can cause it to exist—purely, without help

from minds. (I also agree that there is nothing absurd about, and much to be said for, objective and mind-independent goodness, but that's another story and this paper'd really better not go there. Meanwhile, pretend.) Then if it would be good for  $e$  to help construct  $f$ , this might team up with the funky fact  $\{if\ e\ exists,\ e\ helps\ construct\ f\}$  to cause  $e$ . However spooky, that would be a teleological explanation of  $e$  if anything would be. And if we reached that spooky discovery the first thing we would do is stop trying to demythologize or ignore Aristotle on final causes, and the second thing we would do is consider  $f$  a function of  $e$ , as required by the general theory.

### *The Way of Representation (Desired Means)*

We sometimes speak as if cognitive states (e.g., perceptions, beliefs, hypotheses) are under certain conditions caused by the facts they represent. We also speak as if some facts involving the future (especially possibilities, probabilities, and conditional facts) can exist before the future unfolds. If we are right, then a fact *that  $p$  at time  $t$*  can exist before time  $t$ , and perhaps before time  $t$  it can cause a future-oriented belief (or other cognition) *that  $p$  at time  $t$* . Now substitute the funky fact above for  *$p$  at time  $t$* . Suppose

*$\{if\ e\ exists,\ e\ helps\ construct\ f\}$  constructs a belief that  $\{if\ e\ exists,\ e\ helps\ construct\ f\}$ .*

What would such a belief—underlined above—do in reasoning? The belief has the general form of a means-end belief; it is equivalent to the belief that  $e$  would be a means to  $f$ . So if one had a desire for  $f$ , and had the belief, one might reasonably form a desire for  $e$  as a result, through means-end reasoning. And if one is effective enough in action this desire for  $e$  might lead one to construct  $e$  itself. So when a desire for  $f$  constructs a desire for  $e$  through means-end reasoning, and this constructs  $e$  through effective action, and when the funky fact constructs the relevant means-belief, the funky fact indirectly constructs  $e$ , as follows:

$$\begin{array}{l}
 \text{funky fact} \rightarrow \text{belief in funky fact} \\
 + \qquad \qquad \qquad \rightarrow \text{desire for } e \rightarrow e \\
 \text{desire for } f
 \end{array}$$

This would satisfy the formula in the general constructive theory, so  $f$  would be a function of  $e$  when  $e$  is constructed as a desired (intended, believed-good) means to a desired end  $f$ . When  $e$  doesn't in fact construct  $f$  (even though in a softer sense it "helps" construct  $f$ ),  $e$  malfunctions.

I think this captures what is plausible about Searle's attributional theory of functions. First, Searle places no restriction on whose values generate functions for an entity. He says "our" valuing survival generates a function for my heart of pumping blood, but so long as there is one being in the universe who values my death—a general misanthrope, or a bitter ex, or the neighbor's dog—for all Searle says this equally generates a function for my heart of ceasing to pump blood. And doubtless in this wide world there is already someone who values glorifying God by making thumping noises, so my heart as well as his heart already has that function. But there is no useful sense in which my heart is defective whether it pumps blood or not, and no useful sense in which my heart's thumping is *its own* function. By restricting the role of beliefs



and values to those that help construct an object—the *designer's* or *maintainer's* intentions and beliefs—the modified version of Searle's theory enforces reasonable limitations on whose attributions stick.

Second Searle seems to think that valuing  $f$  makes  $f$  a function of anything that is a means to  $f$ , even if the valuer (is a designer or maintainer who) never intended or thought of the means. He says that even if no one ever heard of hearts, a desire for survival makes a heart have the function of pumping blood because hearts in fact contribute to survival by pumping blood. Now braided hair also contributes to survival under various circumstances, such as when some nearby bad guys happen to be out to kill anyone with unbraided hair. But even if a person's survival is valued by the very person who constructs her hair braids (for show), this doesn't give the hair braids a function of thwarting the bad guys. By restricting the function-endowing role to the specific *means-end attitudes* of the designer or maintainer, the modified attribution theory gets this right.

Third, not even designers and maintainers with relevant means-end thoughts may go hogwild in establishing functions. Even if I design and construct a tool, say a small handtool suitable for extracting pomegranate seeds, I cannot endow it with just any old function, such as establishing world peace, and thereby render it defective. Even if I sincerely believe that it would help construct world peace, and make it for that reason, if this belief is not caused by a (funky) *fact* that it would help construct world peace, my attributions and values generate no matching function.

### *Three Ways of Constitution*

The third through fifth Ways (for a funky fact involving an entity to construct the entity) exploit a sense in which whole entities *trivially* inherit constructive relations (effects, or constitution of further wholes) from their constituents. If a part of Angell Hall helps cast a shadow to the west, then trivially Angell Hall helps cast the shadow. If a part of Angell Hall helps constitute an entrance to Mason Hall, then trivially Angell Hall helps constitute the entrance. (We have to be careful not to imagine that every part of Angell Hall is active in casting the shadow, or that every part of Angell Hall is within the entrance, but still, in a sense, and perhaps by courtesy, Angell Hall helps cast the shadow and helps constitute the entrance. I will omit these qualifiers for ease.) The importance of this inheritance is that it is trivial, that it is without further necessary conditions. If we treat  $e$  as a whole composed of constituents, and inheriting constructive powers from these constituents, we can understand a funky fact *{if  $e$  were to exist,  $e$  would help construct  $f$ }* as reducing to facts about what constituents of  $e$  do, perhaps independently of  $e$ 's existence.

The idea works as follows. Let  $c$  be a constituent of  $e$ . Then if  $c$  helps construct  $f$ , trivially  $e$  helps construct  $f$ . Now instead suppose  $e$  does not exist (yet), but  $c$  (already) does, and helps construct  $f$ . Then the following two conditions can easily hold prior to  $e$ 's existence:

if  $e$  were to exist,  $c$  would be a constituent of  $e$ , and

if  $e$  were to exist,  $c$  would help construct  $f$ .

If both conditions hold, then trivially the funky fact holds:

*{if  $e$  were to exist,  $e$  would help construct  $f$ }*.

Since the funky fact is nothing more than the conjunction of the two conditions, for the funky fact to construct  $e$  all we need is for the two conditions to construct  $e$ . This occurs when  $c$ 's activity ( $f$ ) helps construct  $e$ . Then we have  $e$  existing because *{if  $e$  exists,  $e$  helps construct  $f$ }*, which satisfies the theory of functions.

Well, it does so with one further complication. As soon as  $e$  comes to exist through constituent  $c$ 's activity  $f$ ,  $e$  trivially inherits  $c$ 's activity, so it may be in a sense that  $e$  automatically constructs the  $f$  that helps constructs it. If  $f$  is a function of  $e$ , then in that sense  $e$  automatically fulfills its function. To allow for teleological malfunction,  $e$ 's function must be a *further* performance of  $f$ , beyond the performance of  $f$  that constructs  $e$ .

There are three resulting Ways to satisfy the general theory of functions, because there are three kinds of constituents: spatial parts (the Lego blocks helping to make up Angell Hall), temporal slices (the Angell-Halls-on-a-Graduation-Day helping to make up Angell Hall), and aspects (the greyness helping to make up Angell Hall).

### *The Way of Spatial Parts (Containing Systems)*

Let  $p$  be one of the spatial parts of  $e$ , and substitute it in the reasoning about constituent  $c$  above. Here  $e$  is something larger than  $p$ , something that contains  $p$  as well as other spatial parts. For the existence of the funky fact *{if  $e$  were to exist,  $e$  would help construct  $f$ }* we need these two conditions to hold prior to  $e$ 's existence:

if  $e$  were to exist,  $p$  would be a part of  $e$ , and

if  $e$  were to exist,  $p$  would help construct  $f$ ,

This funky fact constructs  $e$  when these two conditions construct  $e$ . This happens when  $e$ 's organization (out of various smaller spatial parts) is created or sustained by the activity ( $f$ ) of part  $p$ , when what  $p$  does helps bind  $e$  together. In this case,  $e$  has (further) performance of  $f$  as a function.

I think this captures what is plausible, for the generation of functions, about Cummins's emphasis on the organization of containing systems. First, Cummins' is rightly concerned about the following trivialization of his theory:

[N]o matter which effects of something you happen to name, there will be some activity of the containing system to which just those effects contribute, or some condition of the containing system which is maintained with the help of just those effects. Heart activity, for example, keeps the circulatory system from being entirely quiet, and the appendix keeps people vulnerable to appendicitis.

To exclude these nonfunctions, he tries restricting functions to relatively complex systems organized out of relatively simple parts. But this does not do the trick. By his standards a heart's pumping blood has to count as a simple part of the whole relatively complex circulatory system. But then a heart's noisily pumping blood must count as an equally simple part of the whole relatively complex noisy-circulatory system. Furthermore, even simple systems can succeed or fail, in as full-blooded a sense as complex ones can. On the modified organizational view here, what excludes the heart noise is that the noise doesn't help to organize the circulatory system (or the noisy-circulatory system), the noise doesn't help bind or keep the spatial parts together.

Second, under the stated conditions the organization of a system does not necessarily endow the *parts* of the system with functions, but it does necessarily endow the *whole* with functions. If  $e$ 's part  $p$  ceases to perform  $f$ , something is wrong teleologically with  $e$ . It's not necessarily the case that  $p$  malfunctions, since the problem may lie elsewhere in  $e$ . But  $e$  as a whole malfunctions. If a heart doesn't pump blood, this may be because all the blood has drained out through a hole in the foot. That's not a defective heart, it's a defective foot—a defective containing system. Hearts can get their own functions by themselves having parts that organize into hearts, or in any of the other Ways described above and below.

### *The Way of Temporal Slices (Self-driven Survival)*

Now let  $s$  be one of the temporal slices of  $e$ . In this case,  $e$  is a temporally extended “worm”. For  $e$  (the whole extended worm) to exist is for  $e$  to *survive* (or persist) from one time to another. For the existence of the funky fact *{if  $e$  were to exist,  $e$  would help construct  $f$ }* we need these two conditions to hold prior to  $e$ 's survival:

if  $e$  were to survive,  $s$  would be a slice of  $e$ , and

if  $e$  were to survive,  $s$  would help construct  $f$ .

The funky fact constructs  $e$  when these two conditions construct  $e$ . This happens when  $e$ 's survival is caused by  $e$ 's own past activity ( $s$ 's doing  $f$ ).

As before I think this captures what is plausible in Bigelow and Pargetter's propensity theory. As in their theory, there is no need for the activity of “ancestors” of  $e$ . This is because  $s$ , being a time slice, is capable of causing  $f$  and thereby causing  $e$  to survive (persist). You can

think of later time slices as being reproduced from ancestral time slices, in which case the later time slices have  $f$  as their etiological functions, or you can think of them all as being part of the worm, so that the worm itself has nonetiological functions. The worm as a whole needn't have ancestors. Even if you have no ancestors, you can develop functions by contributing to your own survival.

But there should still be a need for *actual* survival activity in the immediate, recent, or distant history of the creature. Otherwise there is no way to apply their notion of a propensity, because as Bigelow and Pargetter acknowledge, propensities are relative to particular conditions:

[A] character [= feature] may confer propensities which are survival enhancing in the creature's usual habitat, but which would be lethal elsewhere. When we speak of the function of a character, therefore, we mean that the character generates propensities that are survival-enhancing in the creature's natural habitat.

For a creature without any history (such as Swampman at the first moment) there is no such thing as its "usual" habitat. Even for a creature with a history, we can't count "usualness" merely by time spent. Suppose that unbeknownst to us, after each five minute period we spend in this dog-eat-dog world, we are whisked away to a protected environment for an hour to "recharge" in suspended animation, and then rereleased into the wild for the next five minutes. Statistically our usual habitat is the recharging station. But our function-relevant habitat is our statistically unusual world where we (even if not our ancestors) perform actual survival-enhancing activities, as reflected in the modified propensity theory.

Given a (possibly very brief) history of survival-enhancing activity in an environment, propensities in that environment drop out as unnecessary. If a creature survives by its own activity in an environment, but more by *fortunate* activity than by the exercise of a reliable propensity, this still establishes a norm for similar activity to live up to. Full-blooded teleological explanation allows not only for entities that happen not to fulfill their functions, but for entities that *cannot* (reliably, without great luck) fulfill their function in their natural habitat—entities without a propensity to fulfill their function. A failing heart does not stop failing at the moment it loses all propensity to pump blood. At that moment it fails worse.

*The Way of Aspects (Reproduction and Selection)*

Finally, let  $a$  be one of the aspects of  $e$ . For the existence of the funky fact *{if  $e$  were to exist,  $e$  would help construct  $f$ }* we need these two conditions to hold prior to  $e$ 's existence:

if  $e$  were to exist,  $a$  would be an aspect of  $e$ , and

if  $e$  were to exist,  $a$  would help construct  $f$ .

The funky fact constructs  $e$  when these two conditions construct  $e$ . Assuming aspects are abstract objects (e.g., properties), and barring direct causation by abstract objects, for an aspect of  $e$  to cause  $e$ , it must be instantiated in something prior to  $e$ . Suppose that  $a$  is so instantiated, and thereby its activity ( $f$ ) constructs  $e$ . The prior thing is similar to  $e$  (at least in sharing aspect  $a$ ), and its similarity is relevant to causing  $e$  (via  $f$ ). This prior thing is therefore an “ancestor” of  $e$  in Millikan’s sense, one that “reproduced” to form the similar “copy”  $e$ . Now, since the ancestor did  $f$  in virtue of having property  $a$ , we can say that  $a$  helped construct  $f$ . Unless  $e$  is very different from the ancestor, then, if  $e$  had existed (perhaps in place of the ancestor), it would have helped

construct  $f$ . So again we have the preexisting funky fact *{if  $e$  exists,  $e$  helps construct  $f$ }*, and we satisfy the general theory of functions.

I think that tracing out this line of thought in more detail would lead to an improved formulation of the etiological theory of functions, better able to handle counterexamples and complaints about it as a stipulative change of subject. But for now suffice it to say that there seems to be a way to justify versions of all the leading theories of functions, under the unified “ $e$  exists because  $e$  would construct  $f$ ” rubric.

### ***Job One is Done; Job Two is Due***

While a history of reproduction and selection stands as one Way to have functions, it is not the only Way. Things suddenly created by the Form of the Good can have functions via their classical final causes. Things suddenly created by intelligent design can have functions by the means-end reasoning of their designers. Things suddenly created by lightning in the swamp can quickly acquire functions by their own survival activity and the activity of their parts in maintaining their organization. Under the right circumstances Swampman can have conscious mental representations from moment two, even if not from moment one. Our assurance that we are conscious, in the face of the possibility that we are Swampfolks, need not rely on our theories of evolution by natural selection of distant ancestors. Though not quite immediate and certain, it is supported by our knowledge of our own continual activities to stay alive and in one piece.



## References and Suggestions for Further Reading

Aladdin's lamp is rubbed in Huxley's *Lessons in Experimental Physiology*, p. 210. The stage-setting quotes about conscious experience are from Nagel's 'What is it like to be a bat?', reprinted in his *Mortal Questions*, and Levine's 'On leaving out what it's like', reprinted in Davies and Humphreys' *Consciousness*.

For panpsychism, see section 8.4 of *The Conscious Mind* by David Chalmers, the most dogged Doubting Thomas. Daniel Dennett is the best source for keeping contentious cases of consciousness contentious—see especially 'Why we can't make a computer that feels pain' and 'Are dreams experiences?' in *Brainstorms*, and his reliance on language and culture in *Consciousness Explained*.

Over the course of several exchanges, Ned Block proves to be the most dogged critic of the idea that consciousness requires meaning, and Michael Tye proves to be the most dogged defender of the idea. See William Lycan's "Representational theories of consciousness" in the online *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/consciousness-representational/>.

Ruth Millikan's *Language, Thought, and Other Biological Categories* best develops the case that meaning depends on function, and best develops the etiological theory of functions. But "Biosemantics", reprinted in her *White Queen Psychology*, provides the easiest introduction. The most challenging counterexamples to etiological theories of functions can be found in Christopher Boorse's "Wright on Functions" (*Philosophical Review*, 1976).

For nonetiological theories, see Bigelow and Pargetter's "Functions" (*Journal of Philosophy*, 1987), Cummins' "Functional Analysis" (*Journal of Philosophy*, 1975), and the first chapter of Searle's *The Construction of Social Reality*. For classical teleology and abstract causation, see Plato's *Republic*, Book VI, 509b, and Andrea Falcon's "Aristotle on Causality" in the *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/aristotle-causality/>. Leslie's *Infinite Minds* (quoted from pp. 160-161) is an impressive defense of an abstract teleological explanation of concrete reality.

Swampman first appeared in Donald Davidson's "Knowing one's own mind", reprinted in his *Subjective, Intersubjective, Objective*. Or maybe Davidson last appeared in Swampman's paper of the same name.