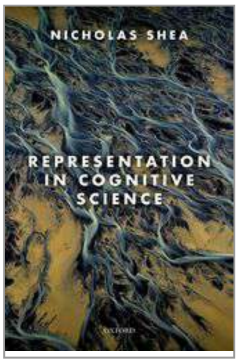

Representation in Cognitive Science
Nicholas Shea

Print publication date: 2018
Print ISBN-13: 9780198812883
Published to Oxford Scholarship Online: October 2018
DOI: 10.1093/oso/9780198812883.001.0001

Standard Objections

Nicholas Shea

DOI:10.1093/oso/9780198812883.003.0006

Abstract and Keywords

Varitel semantics has several resources for dealing with indeterminacy. It gives rise to more determinate contents than informational semantics or consumer-based teleosemantics. The remaining indeterminacy is a virtue: it is what we should expect in simpler systems with fewer interacting components. Some of the non-conceptual representations in our case studies exhibit some features exemplified by concepts: semantically significant constituent structure; unsaturated components; and limited, domain-specific generality. A historical component is needed to bring into view the explanandum to which representational explanation is directed, namely successful and unsuccessful behaviour. We should not expect representational explanation to get a grip, in these simple cases, without some period in which behavioural outcomes have been stabilized. Even a short period of interaction will establish some task functions and contents. Varitel semantics does not reduce misrepresentation to malfunction. Misrepresentation does not imply failure to perform a task function, nor the converse.

Keywords: indeterminacy, qua problem, disjunction problem, teleosemantics, non-conceptual representation, semantic structure, swampman, historical function, norms, naturalism

6.1 Introduction

The positive story is now on the table. We have seen how different accounts of content are suited to dealing with the representations involved in different cases. For each case study, the aim was to give a theory of content that is empirically well-supported and accounts for the way contents are used to explain behaviour. The varitel framework was also designed to produce accounts of content which overcome the most important objections to teleosemantic and other existing theories of content. This chapter will address those challenges explicitly, referring to the existing literature in more detail than when setting out the positive accounts above.

Section 6.2 shows how the approach deals with problems of indeterminacy: distality, disjunction, the qua problem, and so on. The accounts do not deliver perfectly **(p.148)** determinate contents, but I will argue that the level of determinacy achieved is appropriate to the nature of the systems whose behaviour is being explained. Section 6.3 turns to systematicity and productivity, pointing out that the systems we have been discussing do not generally show the kind of compositionality present in natural language sentences. In §6.4 we look at swampman and related challenges to the idea that representational content should depend on a system’s history. Finally, in §6.5 we briefly ask what kind of normativity attaches to representational contents of the kind we have been discussing here and consider the objection that the normativity of representation is categorically different from, and so cannot be based on, the normativity of function (which has not been my aim).

6.2 Indeterminacy

(a) Aspects of the problem

A few examples are standardly used to pose indeterminacy problems for theories of content, most prominently the frog's tongue-dart reflex. I will use that example to illustrate the way my approach deals with various aspects of indeterminacy. Then I will raise and answer those problems for two of our case studies: analogue magnitude representations (in §6.2(b) and (c) below) and cognitive maps (in §6.2(d) below).

I follow the literature in simplifying the frog case. This stylized treatment serves to illustrate the key philosophical issues. So, let us suppose that the frog's tongue dart is triggered by the activity of an array of neurons in the retinal ganglion. Each neuron triggers a tongue dart to a particular location when its activity crosses a threshold. There is mutual inhibition so that only one cell crosses the threshold at any one time.

Consider the putative representation R constituted by the firing of the cell which triggers a tongue dart to location (x, y, z). In typical cases this is caused by a passing fly. Light reflected off the fly and its surroundings passes through the air, through the frog's eye and hits the retina. The pattern of light and shadow hitting the retina excites the retinal ganglion cell R. This causes the tongue to dart out towards the fly at (x, y, z), which is trapped and ingested by the frog. Nutrients from the fly contribute to the frog's survival, and may thereby contribute to the number of offspring it produces. The tongue- dart response is specific to stimuli with tightly delineated characteristics, nevertheless the sensitivity of the system is such that the frog will also snap at little black things that are not flies, like a moving pellet on the end of a fine wire.

The distality problem is to specify which stage of this causal chain contents are concerned with: at the fly or other passing object, or further along the causal chain to the frog's retina and retinal ganglion. At the other end, content could concern proximal effects like the firing of motor neurons or the movement of the tongue, or more distal effects like catching the fly, digesting its nutrients, or eventual continued survival and reproduction.

(p.149) The specificity problem arises for a given stage in this causal chain. For example, a passing fly has many properties. It is a little black thing, a fly (biological taxon), a flying nutritious object (ecological category), something worth eating, something good for the frog, and something that will promote reproductive fitness. R's being tokened increases the probability that every one of these conditions obtains at location (x, y, z). They are not coextensional:[1] R can be set off by little black things that are not flies. Another aspect of specificity arises particularly with teleosemantics, because of its reliance on conditions under which behaviour prompted by R promoted survival and reproduction. That seems to let all sorts of non-specific conditions into the picture: that the prey is not poisonous, that there is no predator nearby that will be alerted to the frog's presence; also background conditions like the presence of air between frog and prey, and normal gravitational forces.

Finally, there is a disjunction problem in that any two or more of these conditions can be put together to produce another condition with which R also correlates. For example, R correlates with: *flying nutritious object at (x,y,z) or little black pellet at (x,y,z)*. In its specific form the disjunction problem is the problem that, for any two worldly conditions C1 and C2 which are candidates for the content of R, C1-or-C2 is a candidate. If R carries correlational information about each condition it will more strongly probabilify their disjunction: $P(C1vC2|R) \geq P(C1|R)$. 'Disjunction problem' is often used more broadly as an umbrella term for all these kinds of indeterminacy. I call them all instances of the problem of determinacy or indeterminacy.

In the past, discussions of indeterminacy have got bogged down in swapping intuitions about what is represented in a given case. Where the representations in question are things like beliefs, desires, and conscious states, we at least have some reason to think that our intuitions about content could get some traction. With the firing in the frog's retinal ganglion cells, we have no such reassurance. There is little reason to give any weight to intuitive judgements about what is represented. The same applies to the case studies we have examined. Instead we have been asking how representations explain behaviour, and what representational contents can underpin those explanations. In Chapter 2 I argued that these explanatory practices are an appropriate constraint on theorizing about content. So, the test of a theory is not that it should deliver intuitive content attributions, but that it should deliver content attributions suited for the explanations of behaviour in which representations figure. Whether or not the contents are appropriately determinate needs to be assessed in that light.

(p.150) (b) Determinacy of task functions

Our first resource is the determinacy of task functions. Recall that task functions are robust outcomes that have a stabilized function or a design function. Out of all the factors that do or could affect the internal processing that produces an outcome F or the consequences that flow from producing F, only a few would be cited in a causal explanation of how F was produced and systematically stabilized by natural selection, learning, or contributing to the persistence of the organism (§3.4d).[2] That is a very substantial restriction on candidate contents.

To apply my framework to the case of the frog, we need to identify task functions and the internal mechanisms that subserve them. The fly-capture mechanism gives the frog a disposition to catch passing flies, which is robust in the face of perturbations and different starting positions. That is, a robust outcome function of the behaviour prompted by R is to trap a fly at (x, y, z). Plausibly, this disposition is the result of selection: past organisms achieving the output of capturing flies is part of the explanation of why there are systems around today with this robust disposition. So catching a fly at (x, y, z) is a task function of R. This task function is achieved synchronically by having an internal mechanism that gathers incoming information, activates one of a range of possible intermediate states (R), and executes a tongue strike in a corresponding direction. That is a simple algorithm, which makes use of the correlation between R and the location of flies on the input side, and between R and the direction of a tongue dart on the output side.[3]

There are a variety of ways of describing the outcome that contributed to selection; for example: catching a fly, a flying nutritious object, or a little black thing. Fodor has argued that, if the categories *fly* and *little black thing* were coextensional in the history of the frog, then considerations based on natural selection cannot choose between them (Fodor 1990, p. 72). That is mistaken. Selection is a causal process. Causal explanation does not in general permit substitution of coextensional properties. Facts about what has been selected for are based on these causal explanations (Godfrey-Smith 1994a, p. 273; 2008)[4]. Only some of the downstream effects of tokening R have been responsible for that disposition's contribution to survival and reproduction. Properties like being small and black do not cause the frog to survive or reproduce (cp. Price 2001, ch. 5, §2). It is because something nutritious was captured that the behavioural disposition was selected. This excludes *little black thing*. It also excludes *something that will promote fitness*, since that is a restatement rather than an explanation of how an outcome leads to survival and reproduction.

Considerations about selection/stabilization are sometimes thought to generate very detailed contents: R represents that there is a flying nutritious object at (x, y, z) **(p.151)** that is not poisonous, contains proteins needed by the frog's physiology, is not moving too fast to be caught, etc. These conditions are of course relevant to whether behaviour prompted by R was stabilized. However, it is a general feature of causal explanations that they do not mention all potentially relevant details, still less the absence of potential defeaters. Explanation has something to do with capturing patterns and generalizing across many events. That counts against picking out events in very fine-grained ways. This is not the place for a general theory of causal explanation, so I rest with the observation that task functions inherit the determinacy of causal explanations of stabilization. That also counts against background conditions figuring in the content, conditions like *fly at (x,y,z) and gravitation is normal*, or *fly at (x,y,z) in air or some similar light-transmission medium*.

This deals with some specificity issues but leaves others open. It does not choose between the following as the type of object to be captured at (x, y, z): fly, flying nutritious object, or object worth eating. We need to look at how task functions converge with correlational information to choose between these contents, at least between some of them (see next subsection).

Let’s see how the determinacy of task functions helps in one of our case studies of UE information (Chapter 4). Consider the analogue magnitude system. It is deployed in situations where behaviour is conditioned on the relative numerosity of collections of objects, doing so by using internal correlates of numerosity and comparing them. We can make the plausible assumption that behaviour involving this system results from reward-based learning, where the rewards arise from acting based on the numerosity of the objects tracked. For example, the animal or person has had the chance to learn that, across a range of different rewarding collections of objects, selecting the more numerous collection leads to higher reward. In animal experiments the reward schedule is devised by the experimenter so we can be confident that numerosity is the basis of reward. It is plausible that some natural learning situations also have that structure, and that learning of this kind, aversive as well as appetitive, underpins behaviour driven by the analogue magnitude system. If we simplify to consider just these comparative situations, the analogue magnitude system is an intermediate in achieving the task function of selecting the more numerous collection of objects. That goes a considerable way to making *numerosity*, rather than other related properties, figure in the contents represented.

(c) Correlations that play an unmediated role in explaining task functions

So, determinacy of content flows in part from task functions not being unduly indeterminate. Further determinacy derives from the requirement that the correlations that are content-constituting are those which play an unmediated role in explaining the system’s ability to achieve its task functions. This calls for a convergence between correlations and task functions. As a result, the appropriate level of distality is constrained by the task function being explained. With the frog, that place is at the location of the fly. The correlation between R and the location of a fly offers an unmediated **(p.152)** explanation of the ability to capture flies at (x, y, z). The correlation between R and a pattern S of light and shadow on the retina could be used to explain fly capture, but less directly: because S correlates with there being a fly at (x, y, z), R’s correlation with S can help explain how the frog manages to catch flies. Even though the correlation between R and S may be tighter, this is a mediated explanation. At the input it is the correlation between R and something distal, at the location of the fly, which plays an unmediated role in explaining the achievement of the task function. At the output end, R correlates with producing the distal result (capturing a fly) which is the task function. When asking why the whole R-involving mechanism was stabilized, the fact that R correlates with catching flies figures unmediatedly in the explanation. The same considerations imply that R does not end up representing all the intermediate links in the causal chain from fly to R to fly capture.

The same reasoning applies to the analogue magnitude system. It is activated on the basis of prior internal states that track individual objects (or events or other entities). So, its states correlate with prior internal states, as well as with patterns of light on the retina, or sound in the ear, etc.; also with causal intermediates between the array of objects and the organism. However, the correlations which directly explain how the organism achieves the task function of selecting the more numerous collection are correlations with a property (numerosity) of the collections being selected amongst.

Some indeterminacies left open by considering task functions and causal explanations of stabilization are resolved when we ask how a collection of correlations carried by a collection of components explains how task functions are performed. We just saw that causal explanations of stabilization might not choose between *fly at (x,y,z)* and *object worth eating at (x,y,z)*. But the frog's fly-capture tongue-dart mechanism is just one of the ways it gets prey. Other internal states correlate with other types of object worth eating to allow the frogs to ingest those. Saying they all just represent *object worth eating* would not capture relevant differences. So, the correlation with flies offers a more perspicuous explanation of how the whole organism achieves its suite of task functions: different mechanisms subserve different tasks and do so in virtue of different correlations.

Varitel semantics does not require that the organism should be able to discriminate the conditions it represents. The frog cannot distinguish flies from little moving black things. Nevertheless, R's UE information concerns flies but not little moving black things. R correlates with flies by being sensitive to sensory features that are a good-enough but imperfect sign of flies, but exploitable correlational information is not restricted to the conditions that a vehicle is the most sensitive or specific sign of. Neither the definition of UE information (§4.2a), nor our evidential test (§4.2c), imply that a stronger correlation trumps a weaker one in constituting content.

(p.153) There has to be convergence between correlation and stabilization, which means contents need to concern conditions that can figure both in causal explanations of stabilization and nomologically based correlations. Perhaps the tongue-dart behaviour in some species of frog has in fact been stabilized in evolution because of trapping just three different fly species S1, S2, and S3 that are prominent in its ecological setting. Then the behaviour was stabilized by catching S1s at (x, y, z), and by catching S2s and (x, y, z), and by catching S3s at (x, y, z). Turning to the correlations carried by R, however, a disjunctive category like S1-or-S2-or-S3 is unlikely to figure in UE information since disjunctive properties are generally poor candidates to figure in causal explanations.[5] The non-disjunctive category fly (the biological taxon), or the ecological category flying nutritious object, look to be better candidates to figure in nomologically based generalizations about what correlates with what. Thus, the need for convergence homes in on more determinate contents than task functions alone would.

Some indeterminacy remains. The biological taxonomic category fly and the ecological category flying nutritious object look to be equally good candidates, both to figure in causal explanations of stabilization, and to figure in the causal underpinnings of exploitable correlational information. So the content of R will be indeterminate between *fly at (x,y,z)* and *flying nutritious object at (x,y,z)*. Furthermore, there is some indeterminacy about the biological taxon *fly*. Is the category restricted to insects (e.g. the order *Diptera*) or should it include other flying invertebrates? Should the biological taxon be understood cladistically (i.e. in terms of shared descent) or in some other way (e.g. in terms of shared phenotypic features or shared DNA)? The content of R is likely to be indeterminate between these options. If we use the term 'flyish' loosely for flying insects, flying invertebrates and flying nutritious objects, then we can say that R represents *something flyish at (x,y,z)*—with the caveat that contents in all of our case studies will be somewhat less determinate than suggested by the very precise tools (i.e. words) we use to express them.

Greater determinacy is achieved in the analogue magnitude case because the mechanism has been stabilized in a wider range of situations. Looking for correlations that are explanatory of comparative choice behaviour across a range of different objects homes in on the correlation with numerosity. That is what the system represents. **(p.154)** Analogue magnitude states correlate somewhat with other features, like total quantity or total surface area of an array of objects, but it is the correlation with numerosity that explains their common role across a range of contexts (as tested in many ingenious experiments). An accumulator system which operates synchronically just like the analogue magnitude system could be present in simpler organisms, and deployed by them in naturally selected behaviours whose acquisition does not depend on learning. If so, those behaviours could have been selected for more specific functions, e.g. to follow the more numerous shoal of fish. If so, the task function serviced would concern something more specific, like the number of conspecifics, rather than numerosity in general. (There are even simpler accumulator systems that do not depend on prior individuation of objects, and simply reflect mass or quantity. Their functions concern quantities but not numerosity.)

Notice that the account does not rely on a representation being caused by what it represents, for example there being a causal connection between the fly and R. It depends only on R carrying correlational information. Suppose R were activated, not by flies directly, but by patches of light on the ground, and that when a patch of light appears, a prey item is likely to land there a short time later. Then P(prey at (x, y, z) | R) would be high, but prey would play no causal role in the tokening of R. The framework would still entail that R represents the location of prey.

This example has the same structure as Paul Pietroski's case of the snorfs and kimus (Pietroski 1992). Kimus are imaginary creatures that are attracted to the red colour of the sun, causing them to climb hills at dusk, thus avoiding their predators the snorfs, who hunt only in the valleys. Pietroski invites the intuition that kimus must be representing redness, rather than something like *snorf-free zone this way*. He argues that the creatures cannot be representing anything about snorfs, since they have no causal sensitivity to snorfs. (The only kimus that have causally interacted with snorfs are historical ones—the kimus that were eaten, hence selected against.) In my view we should give little weight to intuitions about these cases. In any event, our intuitions doubtless draw on imagining a richer picture in which the kimus have conscious sensory experiences and see redness. Once we drop that, the case is wide open. Given my approach to content, all the correlations which R enters into are candidates for content, irrespective of the causal route to tokening R. Correspondingly, if Pietroski's kimus are as simple as the systems in our case studies, they would end up representing the snorf-free direction, even though they have no causal sensitivity to snorfs.

In short, the need for convergence between exploitable correlational information and causal explanations of stabilization is a source of considerable determinacy.

(d) UE structural correspondence

Turning to UE structural correspondence (Chapter 5), the determinacy issues are similar and are answered in a similar way. The cognitive map in the rat hippocampus represents spatial relations between locations. We relied on UE information carried by **(p.155)** place cells to explain route planning, so the convergence we have just discussed between correlation and task function is at work there. Distal correlations with locations figure in an unmediated explanation of task function performance, whereas correlations with sensory features would only offer an indirect explanation. Location is somewhat indeterminate, however, and there is matching indeterminacy in the structural correspondences in play. The co-activation structure on an array of place cells corresponds to absolute locations and the absolute distances between them; to absolute locations and relative distances; to locations picked out relative to some landmarks and their absolute or relative distances; and to locations picked out relative to one another and their absolute or relative distances. There may be general metaphysical reasons why some of these are preferred in a causal explanation of task performance, but that only goes so far. If multiple location- and distance-related features are good candidates for causal explanation in general, then our theory will generate contents that are indeterminate between them.

There is a further, more subtle distinction that we can make with linguistic representations, which may or may not arise with our simpler representations. Place cells act like singular terms, picking out particulars. They could do so indexically, like ‘this’, ‘that’, ‘here’, and ‘now’ in natural language, or non-indexically like ‘London’ or ‘2°W, 10°S’. Individual place cells are clearly saying something more than *I am here now*, since they are reused offline with the same content when calculating shortest routes. But we can ask whether the array of place cells is picking out an array of locations indexically, as something like the locations around here now, or non-indexically, with singular terms that work like names for locations. I can think of three possible answers. The first is that there is a general answer about all these kinds of simple systems; for example, that none of the representations is indexical (or conceivably that they all are). The second answer is that cognitive maps represent in a way that is indeterminate between indexical and non-indexical representational contents. Or thirdly, it may be that the question itself is ill-posed, when asked about a system that does not support a distinction between different ways of picking out its referents. I remain neutral between these answers, accepting that this may be a source of indeterminacy in our case studies.

(e) Natural properties

Since content in these cases is fixed by reference to causal explanations, natural properties will be better candidates. This makes some disjunctive properties unsuited to figure in the content. Arbitrary disjunctions are not good candidates to feature in causal explanations.

This consideration also resists the objection based on ‘reduced content’ (Peacocke 1992, pp. 129–32). R correlates with there being a fly at-(x, y, z)-and-within-the-organism’s-lightcone. That condition certainly applied on all occasions when ancestor frogs interacted with flies in their selection history. However, causal explanations do **(p.156)** not in general appeal to these kinds of ‘reduced’ properties. To give a general characterization of the kinds of properties that are candidates to figure in causal explanations would take us beyond the scope of the present enquiry. It suffices to note here that it is facts about causal explanation that rule out reduced contents.

These points only apply to the kinds of simple systems we are considering here. It is clear that more esoteric contents are not ruled out in more sophisticated representational systems, like human conceptual representation. We persons can represent proximal properties as well as distal properties, contents like *fly and within my lightcone*, and disjunctive contents. Those abilities depend upon the greater complexity of our representational apparatus, especially the combinatorial power of concepts.

(f) Different contents for different vehicles

A final factor at work here is the soft constraint that different representational vehicles should have different contents. That is not an explicit part of what it takes for a correlation to amount to UE information, but it follows in many cases. UE information focuses in on correlational information that is exploited in order to perform a task function. Different vehicles have different effects on downstream processing, so ascribing the same contents to a whole range of different vehicles could miss out on important aspects of the way the system performs task functions, hence would be less explanatory.

For example, suppose we treated all retinal ganglion cells in the frog as having the same content. They carry the information that there is a fly somewhere nearby and trigger catching behaviour. Getting a fly is arguably a task function of all the tongue- dart responses. Retinal ganglion cells carrying information about flies does help to explain how that outcome is achieved. But there are also more specific task functions that go with more specific responses: the function of catching a fly at (x, y, z) is a task function of the response prompted by a particular ganglion cell R. The correlation of R with the coarse-grained condition *there is a fly nearby* could be partly explanatory of achieving that function, but the correlation with *there is a fly at (x,y,z)* is more explanatory. So, the latter gives the content.

Millikan has a similar requirement. Built into her idea of 'most proximate Normal explanation' and 'derived adapted proper function' is the idea that different representations, acted on differently by the consumer, should have different contents (Millikan 1984, pp. 44–5, 97). In my case the requirement is not that every representation within an organism should have a different content. But when there is a stage of processing that admits of a range of mutually incompatible vehicles whose differences make a difference to downstream processing, an explanation of how that processing contributes to performing task functions will generally point to correlational information that is different across those vehicles.[6]

(p.157)

> *Soft Constraint: Different Contents for Different Vehicles*
>
> When a stage of processing can adopt a range of mutually incompatible states R_i, each affecting downstream processing in a different way, correlational information which is different for each of the R_i will generally be a better candidate to be UE information, other things being equal.

In the frog case, it follows that the different retinal ganglion cells represent flies at different locations, rather than all simply representing something like *fly* *nearby*. In the analogue magnitude case it follows that numerosity is being represented rather than something more coarse-grained like *many* and *few*.

The soft constraint applies to mutually incompatible representational vehicles. There is also the question of whether different components within an overall computational process, elements that can be tokened at the same time, can carry the same content. That does arise, for example the visual system contains multiple representations of the location of an observed object. The soft constraint does not rule out such cases. Nevertheless, if we want to see how internal processing carries out computations that are suited to performing task functions, that will generally require different steps to carry different contents.[7] So there are general explanatory reasons that somewhat count against different elements carrying the same content, without ruling it out in a suitably articulated system.

(g) The appropriate amount of determinacy

A final consideration is to ask what the appropriate amount of determinacy is. In these simple cases quite a high degree of indeterminacy may be expected. Lacking many of the moving components of richer representational systems like those found in human belief-desire psychology, it should be no surprise that lower-level systems have more indeterminate contents. In systems with more components those components will often be playing more specialized roles.

In giving representational explanations we are appealing to relational properties of component parts in order to explain the system's behaviour. Components will often stand in a family of closely related relations to a family of closely related distal properties. In the frog, these include the taxonomic biological category *flying insect* and the more physiological category *flying nutritious object*. There is no reason to expect this simple system to support a distinction between representing *flying insects* and *flying nutritious objects*. That is a kind of indeterminacy that flows from the limited complexity of the system.

How best to capture this indeterminacy? One approach is to say that the system carries each of these closely related rival contents, and that we can appeal to any of them in explaining its behaviour. Alternatively, it could be that there is a single natural **(p.158)** property in the vicinity of both candidates that figures in the content, but that we are unable to pick it out exactly, because the language we use is unsuited for doing so, being too precise.[8] On the second option content is not strictly indeterminate, but it consists of a determinate success condition that can only be picked out approximately or disjunctively using the tools of natural language. I don't propose to arbitrate between these options. I rest with the claim that the indeterminacy that remains at this level is unobjectionable.

We noted that less indeterminacy is likely to arise in systems with multiple interacting components. There is also a distinction to be made between indeterminacy at the level of an individual vehicle and indeterminacy at the level of the whole system. This is best illustrated with an example. Recall the system in the prefrontal cortex for deciding the preponderant direction of motion of visual stimuli in one context and the preponderant colour in another context. We saw in §4.6b that applying varitel semantics to this system leaves some residual indeterminacy. The content of the input representation for colour, R_1, is indeterminate between (a) *the majority of dots are red*, and (b) *the colour density is predominantly red*. There is corresponding indeterminacy in the representation C_1 that registers context: between (a) *reward will be based on the colour of the majority of dots on the screen*, and (b) *reward will be based on the predominant colour density on the screen*. However, to be explanatory, a correlation carried by R_1 needs to go with a correlation carried by C_1: (a) with (a) or (b) with (b). There is one set of exploitable correlational information carried by the whole collection of components which includes the two (a) clauses as UE information. There is a second set which includes the two (b) clauses. A disjunctive assignment of (a)-or-(b) to R_1 at the same time as (a)-or-(b) to C_1 will not be UE information. Putting R_1's registering (a) together with C_1's registering (b) is a poor explanation of why the system makes the choice it does. In any event, disjunctive conditions are poor candidates to be exploitable correlational information in the first place (§6.2e above).

So, there are indeterminacies about the overall UE information carried by a system that are not simply recapitulated, component-by-component. Furthermore, the need for UE information to align between components, so that interactions between components make sense in the light of their contents, is a significant constraint on indeterminacy in systems with multiple interacting components. These are both reasons why the residual indeterminacy implied by varitel semantics varies with the complexity of the system in question. That is an appropriate result.

(h) Comparison to other theories

My approach to indeterminacy adopts many of the elements relied on by Millikan's teleosemantics (Millikan 1984, 1989, 1990, 1995, 2004). Contents for Millikan derive from the 'most proximal Normal explanation' of how behaviour prompted by a **(p.159)** representation led to survival and reproduction. Directive content is the output specific to a representation that features in such an explanation. Descriptive content is the condition, specific to a representation, which explains how those outputs led systematically to survival and reproduction. My own focus is on unmediated explanation of the performance and stabilization of task functions. This may cover a wider range of systems, but retains the merits of Millikan's view: indeterminacy is constrained since causal explanation does not generally allow substitution of coextensional properties *salva veritate*; and also by setting aside mediated causal explanations of stabilization. This makes non-natural or disjunctive properties poor candidates for content for Millikan (1990, p. 334), but as with my account, indeterminacies remain as between properties that have equivalent causal-explanatory significance (Godfrey-Smith 1994a, p. 274).

My requirement for convergence between correlational information carried and task function performed is an additional source of constraint (§6.2c and Shea 2007b, cf. Millikan 2009). I am also perhaps more explicit about the requirement that different representations in the same range should have different contents, and about why that is so (§6.2f). Since I do not attempt to apply my account to conceptual representations or conscious states, I have an argument that the indeterminacies which remain are an attractive feature of the account, rather than a failing (§6.2g). Furthermore, as we saw in Chapter 4, giving up the consumer requirement allows us to deal with systems with multiple interacting components—which thereby have relatively determinate contents—more easily.

Papineau also advances a consumer-based teleosemantic theory. For him the theory applies in the first instance to belief-desire psychology. He argues that desires have determinate contents and act as consumer systems for beliefs, which inherit that determinacy. He used to think that, outside the belief-desire system, teleosemantics results in considerable indeterminacy because multiple systems are equally good candidates to count as consumers (Papineau 2003). He now thinks an idea of Neander's can solve that problem (Neander 1995). A component in a system will indeed have many different nested functions (derived from evolution and/or learning), but teleosemantics should only appeal to its specific function, outputs that it produces on its own in the lowest level description in which it appears as an unanalysed part. This leads to a view in which malfunctions only arise from the failings of the component itself, not from interactions with other components (Papineau 2016). My view goes in a somewhat different direction here, as we will see in a moment.

I follow Price in thinking that the way representational properties feature in the explanation of behaviour should help us to characterize their nature (Price 2001, ch. 4, cp. my desideratum §2.2), also in requiring representations to carry correlational information (for me, in one class of cases). Price adopts Neander's useful distinction between 'high church' and 'low church' teleosemantics (Neander 1995). High church teleosemantics ties content to explanation of the success of behaviour prompted by a representation. Low church teleosemantics focuses on the way representations are produced and ties **(p.160)** content to the actual discriminative capacities of the organism. Pietroski's argument about the snorfs and the kimus was a push in the low church direction. Millikan and Papineau argue for the high church view. Dretske (1988) and Ryder (2004) are also in the high church, since they tie content to properties that explain successful behaviour.

Price herself adopts a high church view. She argues that teleological considerations, supplemented with some plausible principles, can deliver determinate contents (2001, ch. 3). Price's immediacy and abstractness conditions have a similar effect to my focus on correlations that enter into an unmediated explanation of how the system performs its task functions. Like Papineau, Price relies on Neander's idea that the relevant functions of a device are things that it can do by itself (in servicing a wider mechanism, or 'governor'). My approach has a rough analogue of this when there are multiple components since I ask what each component contributes to an algorithm realized in the system so as to perform its task functions (Chapter 4). However, my task functions are decidedly not limited to outputs for which a single component is responsible. They are outputs of the whole organism and depend on interactions amongst its components. Nor do I think a vehicle is only misrepresenting when something goes wrong with the component responsible for producing that vehicle. Many cases of misrepresentation are caused by malfunction in upstream components; and even when all the internal processing is operating as it was designed to, misrepresentation can occur when the environment is uncooperative (i.e. unlike it was during stabilization). (I also differ from Price in giving up on the need for a consumer, and in the pluralism that allows for different kinds of exploitable relations and different kinds of functions.)

Karen Neander is the leading proponent of low church teleosemantics (Neander 1995, 2006, 2017). She argues that content concerns the objects and properties an organism is causally sensitive to and should be tied to conditions that it can discriminate between. One argument is based on the idea that a component has not itself malfunctioned if the external environment is uncooperative (Neander 1995). So, for example, if a frog snaps at a little black thing that's not a fly, that should not be counted an error, because there is no malfunction within the detection mechanism.[9] But I have argued that facts about how components of a system interact with one another are not enough to get content explanation off the ground (§2.3). We need to look at how they are designed to interact with the distal environment. Long-armed functions can go wrong when the environment is uncooperative without that being attributable to the failure of any of the internal workings.

A second argument is based on 'response functions' and a detailed case study of the science of prey capture in the toad (Neander 2006, 2017). Neander rightly observes that scientists have been concerned to discover how the toad manages to track prey. That is a different explanandum (2017, p. 119). I link content to explanation of behaviour. The scientists are trying to work out how the toad manages to track—I say represent—prey in its environment accurately enough to survive (2017, p. 108). I don't see why long-armed etiological functions need be tied to discriminative capacities, and I certainly **(p.161)** reject the verificationist claim that an organism with non-conceptual representations can only represent what it is capable of discriminating (2017, p. 120). We might well be interested in how an organism manages to discriminate the things that it represents, but to formulate that question we need to leave room for a gap between the things it represents and the way it discriminates those things. Verificationist contents are poor at explaining unsuccessful behaviour—for example, explaining why things go badly for a toad when it moves into an environment rich in little black moving things that are not flies. Basing content on discriminative capacities also means that Neander has to add a special purpose principle to make contents come out as distal (2017, p. 222).

Does my account entail, then, that organisms will never represent perceptual features like being a little black thing and will only ever represent properties like *fly*? Surely the human perceptual system represents features of an object, like its size, shape and velocity, on the way to categorizing it as a *fly*? My account agrees. A suitably articulated system does end up representing more sensorily specific features of an object on the way to representing it under more general categories. We saw that for the visual system in §4.7, where my account delivered representations of chromatic properties and local motion properties. That flows from applying the varitel framework to a system in which information-processing is broken down into multiple interacting components; especially when, as in the human perceptual system, a single perceptual representation feeds into many different kinds of downstream processing and behaviour. So, on my view contents can concern perceptual features of objects, and perceptual systems in complex organisms will typically represent features which they then use to track behaviourally significant categories of object. None of that is found in the toad's simple prey-capture mechanism, at least in the stylized version described here.

Other authors have different proposals about which properties are good candidates for representational content. Ryder works this out with respect to a particular mechanism, SINBAD, whose function is to detect statistical regularities in patterns of input (Ryder 2004). As a result, SINBAD's states end up referring to properties that explain those regularities. Martínez makes an ontologically more committed version of a similar move (Martínez 2013). He argues that homeostatic property clusters are privileged candidates to figure in representational content.[10] Artiga generalizes that view: content is given by a subset of properties which explain why candidate properties tend to co-occur, even when there is no homeostatic property cluster (Artiga in submission).

The problem with all three of these proposals is that they focus on the way the information that the system is responding to is generated: the homeostatic property cluster (if there is one) that underlies the incoming information, or the source of statistical dependencies amongst sources of information. This property need not be the same as the property or properties that constitute and explain successful behaviour. Speaking loosely, an organism does not care what the most informative property is; it cares about what needs to be in place for its behaviour to be successful. For example, consider a rainforest frog that spawns in small pools of water which, in its habitat, are almost all **(p.162)** found in *Nepenthes* pitcher plants. The frog recognizes spawning locations by detecting the sight, smell, and typical locations of *Nepenthes* plants. The property underlying this regular statistical structure is the presence of the genus *Nepenthes*. However, the success of its spawning behaviour just turns on finding a suitable pool of water. Spawning in a pool that happened not to be in a pitcher plant would not count as a failure. My theory implies that the frog is representing the location of water rather than the location of *Nepenthes* plants.

In short, while taking much inspiration from earlier teleosemantic treatments, my account departs from them in important ways.

6.3 Compositionality and Non-Conceptual Representation

An important feature of the representations found in the human belief-desire system is that they make use of concepts. Concepts are reusable elements which do not make claims or set goals taken individually: they are unsaturated. Only when put together do they form a saturated representation with a complete correctness condition or satisfaction condition. This book does not attempt to deal with concepts and how they get their content. Concepts do, however, have several features which are also found in some of our case studies: semantically significant constituent structure, unsaturated components and (limited) generality.

I reserve ‘concept’ for the unsaturated personal-level representations that are expressed in language and combine to form beliefs and desires.[11] ‘Non-conceptual’ covers all representations that not are concepts or constructed out of concepts. Therefore, all the representations in our case studies are non-conceptual, although as we will see some share some features of conceptual representations.

Concepts obey a wide-ranging generality constraint: they can be recombined liberally with other concepts in the thinker’s repertoire. For example, any one-place predicative concept, F, can be combined with any singular concept, a, to produce a saturated representation, Fa, which is a candidate for belief. If the thinker can also think Gb, then they have four reusable components, so for example they automatically have the capacity to think Fb.

I will use the term ‘saturated’ so as to include non-conceptual representations with a complete correctness condition or satisfaction condition, whether or not they are constructed out of unsaturated elements. So, for example, an output node in the simple feedforward connectionist network in §4.3 is a saturated non-conceptual representation even though it has no semantically significant constituent structure (its complete correctness condition is: *the object encountered is in category A*).

(p.163) Predication is involved when unsaturated concepts are put together to form a saturated representation. Predication is absent from most of our case studies, with the exception of offline use in the rat hippocampus. However, many of the case studies do exhibit semantically significant constituent structure of a simpler kind. They also exhibit some local recombinability, and thus some limited generality. None meets the kind of wide-ranging generality constraint met by concepts.

Consider the visual system which detects plaid motion (§4.7). One layer represents chromatic properties at locations in the visual field, another layer represents motion direction at locations. This gives the system a limited kind of systematicity: for each location, it can represent that location as having a range of colours and it can represent that location as having a range of motion directions. But this is the systematicity of a list. There are no a singular terms representing locations and nothing acts as a recombinable representational constituent. Nor are the colour and motion representations tied to the same vehicle. If the vehicles representing motion direction in one part of space were selectively lesioned, the system would retain the capacity to represent colours at those locations. Each layer independently forms saturated representations about colour and motion respectively.

Now consider the case in §4.6: a single distributed representation in prefrontal cortex (PFC) represents both the colour and the average direction of motion of an array of moving dots. This vehicle represents colour and motion at the same time. The well-known bee dance example is similar: a single dance represents both the direction and the distance of a source of nectar. In both cases the system exemplifies a limited form of systematicity. A range of direction representations can be combined with a range of distance representations. But they do not involve anything like predication. There are no unsaturated components, that contribute to the semantic value but fail to have a correctness condition on their own. If the dimension that goes with colour is removed in the PFC case (as it is effectively, in direction choice trials), the remaining dimension still represents that the array is moving in a certain direction. If the number of waggles were indistinct or ignored, a bee dance would still represent the direction of a nectar source. Each dimension is acting like an independent saturated representation with a complete correctness condition.

An important question to ask is whether a representation has semantically significant constituent structure. The plaid-motion system has two different representational vehicles, neither of which has semantically significant constituent structure. The PFC colour-motion system has a single vehicle with two semantically significant dimensions of variation. That is semantically significant constituent structure. There are a range of syntactic states each of which can represent both the colour and the average direction of motion of a stimulus. The correctness condition for these representations is something like: *the currently presented array is of colour abc and is moving in direction r*. Two elements of that correctness condition correspond to two dimensions of variation of the vehicle (colour and motion direction). Other elements of the correctness **(p.164)** condition do not correspond to any dimension of variation in the representational vehicle: e.g. which stimulus bears these properties and when.

The ability to vary two features independently is an important kind of semantically significant structure and a notable source of the representational power of that system in the PFC. But it is important to distinguish this from having unsaturated constituents. In the PFC case and the bee dance case, neither dimension of variation is predicated of the other. Each on its own is capable of making a saturated claim. The human conceptual system by contrast makes use of unsaturated elements and predication.

Unsaturated elements can arise when there are multiple dimensions of variation corresponding to different features and there is no stabilization story to be told about how each could prompt behaviour independently. Conditions for successful behaviour only arise when all are tokened together. (Perhaps no behaviour is produced at all when one dimension of the vehicle is tokened on its own, or such behaviour as is produced played no role in stabilizing the mechanism.) Offline place cell activation may be like this (§5.7b). Nothing functional follows from the activation of a single place cell offline in isolation. Co-activation of two or more place cells is required for the offline system to contribute towards the system’s task functions. In such cases neither vehicle, tokened on its own, has a complete correctness condition. It is only when two place cells are active that the relation of co-activation is instantiated. The activation of the two place cells then forms a representation with a complete correctness condition (e.g. *location 1 is near location 2*). One explanation for this is that offline place cell activation is unsaturated: each cell contributes a location, and only co-activation has a complete content. (I want to remain cautious about whether this is predication in the sense in which sentences involve predication, or whether it is a different way in which components can be unsaturated—a different kind of function application.[12]) A second explanation is that offline activation of a place cell has suppositional content. It says something like *suppose you were at location 1*. I explore that idea further in the next chapter, when we look at descriptive, directive, and other modes of representing (§7.5b).

The third notable feature of the human conceptual system is its compositionality: any of the unsaturated representations can be combined with any other of the right kind to form a saturated representation. That is, the compositionality of representational vehicles leads the system to obey a wide-ranging generality constraint (Evans 1982), i.e. to exhibit systematicity (Fodor 1987b). The honeybee nectar dance has two semantically significant dimensions (corresponding to distance and direction). Any value of one can be combined with any value of the other. This is only a very limited kind of systematicity. The PFC colour-motion case also displays this kind of limited, domain-specific systematicity. The distributed PFC representation can combine any **(p.165)** claim about colour with any claim about motion. Furthermore, given its flexibility, the PFC is likely to have the capacity to also to represent other objects, properties, features, and events. But this is not the wide-ranging systematicity of concepts, where any representation can be put together with any other. The cognitive map and even the PFC system meet only a limited, domain-specific generality constraint. But this is a step in the direction of the full-blown generality constraint obeyed by concepts.

Millikan claims that time and place of production are constituents of simple signs like the honeybee nectar dance. Time and place may figure in the correctness condition, but as with some other elements in the correctness or satisfaction condition of a non-conceptual representation, there is no syntactic type in my sense which corresponds to time or place. For example, there is no singular term picking out location in the way activation of a place cell picks out a location in the rat’s cognitive map. Variations in a vehicle are semantically significant when they can take a range of values and their variation makes a difference to downstream processing and/or behaviour. The representational theory of mind is based on the idea that aspects of a vehicle are being exploited for the relation they stand in to features of the environment. Where the mechanism is not able to do different things for different times of tokening, but operates in just the same way whenever the vehicle is tokened (as in the PFC and bee dance cases), time of tokening is not a semantically significant aspect of the representation. Indeed, it is hard to see how time of production could be causally effective in downstream processing unless it is marked or measured in some way.

Another important feature of these cases also sometimes gets the label ‘systematic’, which is that they form an organized sign system (§5.5; Godfrey-Smith 2017, p. 279). There is a straightforward mathematical relationship between a dimension of variation of the vehicle and the content represented. More activation along one dimension represents a greater amount of motion. Similarly, there is a straightforward mapping from direction of the bee dance to the direction of nectar. Learning or evolution produced a mechanism that follows the mapping. As a result, intermediate values, which may never have been exemplified during learning or evolution, come to have appropriate contents. That, too, could fairly be called a kind of systematicity: there are facts about the mechanism as a whole from which it follows that novel representational vehicles carry determinate contents. Having a single mechanism that can respond to a range of cases in a systematic way is also doubtless an advantage for the system.

However, the phenomenon of intermediate values having contents is importantly different from the power that comes from the ability to recombine different representational components, particularly the power of being able to do so in a very general way, as with human concepts.

Representational content derives in part from the situations in which a representation is formed and then acted on to produce behaviour. We saw that the way analogue magnitude representations are produced and used by multiple systems gives them considerable determinacy and establishes reference to numerosity rather than other closely related properties. When we get to concepts, we have syntactic items that are **(p.166)** reused across a wide range of situations as they are combined with other concepts. This gives a wide range of uses involved in fixing their contents—giving them scope to have more specific contents. For example, the concepts HOPE and WANT are used across a wide range of circumstances, for understanding others' behaviour and planning one's own. That is part of what allows them to refer to different but closely related psychological properties. The representations in our case studies are not deployed across such a wide range of uses and so are likely to have less determinate contents than conceptual representations have.

To recap, I have picked out three features of conceptual representations and shown how they each occur in some way in some of the non-conceptual representations in our case studies: semantically significant constituent structure, unsaturated constituents, and (limited) generality. The simple feedforward connectionist system (§4.3) and our simplified version of the visual mechanism for detecting plaid motion (§4.7) involve only representations without semantically significant structure. In both cases the system can token more than one representation at once, but these are separate vehicles—what I have called the systematicity of the list. The PFC colour-motion system (§4.6b) and the honeybee nectar dance exhibit semantically significant constituent structure. A single representation has two independent dimensions of variation, each a saturated representation with a complete content. They do not have unsaturated constituents. When place cells are used offline to calculate shortest routes they arguably function as unsaturated constituents, combining so that their co-activation represents spatial proximity. Finally, the systematicity involved in the place cell, PFC colour-motion and honeybee nectar dance representations means that each exhibits some limited domain-specific generality. None of these systems obeys the kind of wide-ranging generality constraint met by concepts.

6.4 Objection to Relying on (Historical) Functions

(a) Swampman

Perhaps the most prominent objection to teleosemantic theories of content targets their defining characteristic: relying in part on etiological functions to fix content. Etiological functions depend on history: a history of selection, learning, or other interaction with the environment. My accounts of content face this challenge because task functions depend partly on history, and task functions play a role in fixing content.

The challenge is made vivid in the literature by considering a ‘swampman’—an intrinsic duplicate of a human, but one who arises by complete chance as a result of lightning striking a swamp. Swampman would look and behave like a person with mental states, but any theory of content relying on a historical notion of function will imply that he does not have mental representations, states with content, at least at the moment of creation. Where task functions are based on natural selection, only a system that is the result of selection will thereby have content; where task functions are **(p.167)** based on learning or contribution to persistence, a swamp system will not have content until it has undergone some interaction with its environment involving learning or helping the organism to persist. Chapter 3 set out that consequence, illustrated with a toy example (§3.6). This section (i) offers a positive argument that this is the right approach, and (ii) compares my response to others in the literature. I will leave aside task functions based on deliberate design, but they too require history: that a system has been designed or co-opted for certain functions.

We could imagine a swamp system that is an intrinsic duplicate of any of the cases set out in Chapters 4 or 5. The swamp system would have the same behavioural dispositions, so would have robust outcome functions. For example, a swamp duplicate of the system in §4.7 would have a disposition robustly to catch a partly obscured moving object producing plaid motion. It would do so making use of a structure of internal processing, where those internal elements stand in appropriate exploitable relations to distal features of the environment. Since there are robust outcomes involving distal objects and properties, which proceed via a multitude of different proximal routes, there will be distal-involving real patterns in the way the object would interact with its environment, patterns that do not depend on history (which is what I relied on in §3.6). If content did not depend on stabilized functions, but only on the robust outcome function aspect of task functions, then content would inhere in the swamp system. Why isn’t that a perfectly good notion of content?

Recall the distinctive ‘explanatory grammar’ of representational explanation: correct representation explains successful behaviour and misrepresentation explains failure. It is because the success or failure of actions does not depend just on intrinsic properties of the organism or its bodily movements that I argued that this explanandum called for explanation by relational properties of the system (here, relational properties of internal components of the system). What I want to argue now is that, without appeal to history, in the simple cases we are considering here, there are no other ingredients to draw on to make it the case that some consequences should count as successes and others not. That is, the thing which contents are called on to explain—success and failure of behaviour—

is absent in a simple system that lacks history (cp. the ‘no explanandum’ argument, §1.5).

Consider a swamp system corresponding to the plaid motion object catcher in §4.7—call it ‘Catcher’. Compare Catcher to another swamp system that happens to have the robust disposition to reach out just to the outside of the direction of motion of a moving object, so that the object bounces off the edge of its hand and passes by; call it ‘Misser’. Misser robustly achieves the outcome of glancing the edge of its hand off passing objects, and will do that from many starting positions, adjusting in real time for perturbations in the path of the object. If content were founded just on robust outcome function plus appropriate internal mechanism, both Catcher and Misser would have content. Catching the object would count as a success for Catcher, and were it occasionally to miss, that would count as a failure. The converse is true for Misser—the occasional catch would count as a failure.

(p.168) An appropriate tweak of internal workings would turn Catcher into Misser. Suppose that too happens by chance. If Tweaked-Catcher now interacts with an object and drops it, is it successfully achieving the same robust outcome function as Misser, or is it misrepresenting the trajectory of the object and so failing to achieve the robust outcome function it had before it suffered the tweak? If content were founded just on robust outcome functions, then whatever the system is disposed to achieve robustly would count as success. So, Tweaked-Catcher would not be misrepresenting, but would be successfully performing the same robust outcome function as Misser. A swamp duplicate of a human that happened to be disposed robustly to pick and eat a type of berry which is poisonous would count as behaving successfully, even if it would soon die or learn to avoid the fruit. We want our theory to allow that there are cases where error leads a system to pursue a poor outcome robustly; for example, a guided missile that systematically misrepresents its location and so robustly arrives a kilometre north of its target. There is no room for such cases if content is based just on current robust outcome functions. That notion of function does not furnish the resources to constitute some robustly produced outcomes as genuine successes and others as failures.

An approach that builds history into the notion of function can make this distinction. Task functions are established by the convergence of stabilized function with robust outcome function at the time of stabilization. If damage or a tweak to the system alters its robust outcome dispositions, then it will be robustly disposed to produce unsuccessful outcomes—outcomes which are not amongst the task functions of the system. Indeed, it is a strain to apply the notion of success just on the basis of robust outcome function. For a moderately complex system like the one in §4.7, very many different outcomes could be robustly produced through small changes to the internal operation of the system. With such a wide array of outcomes potentially counting, it seems tendentious to label each a potential way of distinguishing successful from unsuccessful behaviour. Too many behaviours potentially count as successful. What is missing is any connection with doing good for the system—the sense that successful behaviours are ones that are or have been beneficial. It is the historically based notion of stabilized function that makes success, as constituted by task functions, retain a connection with goodness or benefit.

That is just an intuition, but it mirrors the argument in Chapter 3 which was based on the underlying motivation for representationalism. Representations get their explanatory bite in these simple systems because there is a real cluster in nature where selection, learning, and contribution to an organism's persistence go along with having dispositions to produce certain outcomes robustly, and with doing so by having internal processing that exploits relational properties of internal components. Severing the connection between function and some kind of consequence etiology takes us outside that cluster. Similarly, a forward-looking notion of benefit or consequence is not one of the items that found the cluster. (Recall also the positive arguments against forward-looking accounts offered in §3.4d and §3.7.) It is the existence of a consequence etiology in the past, even the very recent past, which goes along with producing certain outcomes **(p.169)** robustly—not the fact that those outcomes would (or might) lead to good consequences for the system in the future.

That is an argument for keeping consequence etiology, hence history, in the picture. The kind of history that counts may be very recent. In most of our case studies the stabilized function is based on a history of learning, and is not derivative from evolutionary history. As soon as a swamp system starts interacting with its environment and learning, it will rapidly acquire task functions. So, it won't be long before there is a basis for counting some outcomes as successful and others as unsuccessful, and then we can start explaining the success and failure of its behaviour in terms of correct and incorrect representation.

Similarly, a swamp human will start with only as-if memories but will soon acquire genuine memories of its interaction with the swamp. It will start by having only an empty simulacrum of relations with other people, but will soon start building up friendships with the people it interacts with. The swamp human is importantly disanalogous to swamp versions of our simple systems, since the extra sophistication of its cognitive apparatus, and/or the fact that it is conscious, may make it a genuine representer from the moment of creation. But the analogy serves to illustrate that it is not unusual that mental properties should depend on interaction with the environment and build up very quickly in a swamp system. When we turn to our case studies, like the reaching system in §4.7, a small amount of interaction with falling objects, with feedback serving to fine-tune the system's dispositions, would be enough to constitute catching objects as a task function of the system. So, it wouldn't be long before a tweak to the system which turned a Catcher into a Misser would count as a failing, with unsuccessful behaviour in the latter rightly blamed on its disposition systematically to misrepresent the location or trajectory of objects in its environment.

In short, my answer to the swampman challenge—the objection to content being partly historically determined—is to cut down the scope of the objection and then accept the consequence, offering a positive argument that, in these simple systems, content should depend partly on history. The scope of the objection is curtailed in two ways. First, because my view does not imply that a swamp human would lack contentful conscious states or thoughts. Content of personal-level representational states may indeed be fixed ahistorically. Secondly, because even in our simple case studies, some contents would soon be established once the system has had a chance to interact with its environment.

(b) Comparison to Millikan and Papineau

My answer to the swampman challenge is different from those previously given by Millikan and Papineau, but not radically so. Millikan argues that a swamp creature would not be part of a real kind, and so generalizations from real humans to swamp creatures would be unsupported (Millikan 1984, pp. 93–4; 1996). They could at best be right by accident. Humans are part of a real kind, but that is a historical kind, the species *Homo sapiens*. According to popular cladistic views of biological classification, **(p.170)** species are constituted by shared descent, not by any current property of a population of organisms like shared DNA.

The trouble with this response is that it does not tell us why generalizations about content should go with the historical kind *Homo sapiens*. Human organisms are also physical objects, and as such they obey the laws of gravity. Those generalizations apply to them on the basis of membership of a currently constituted category. So generalizations about how a human falls from a cliff would unproblematically extend to swampman. Why should content properties go in the historical camp? That is particularly puzzling given that swampman looks and talks as if he is susceptible to non-historical generalizations based on attributing representational states in standard ways. And it is agreed on all sides that expectations about behaviour, so formed, would be fulfilled with swampman just as they would be for his real human doppelganger.

Varitel semantics gives us the resources to explain why the generalizations that found content do not extend to swamp duplicates of the systems in our case studies. Duplicates do not fall into the pattern whereby robustness of outcome goes together with internal workings, exploitable relations, and a history of selection, learning, or contribution to an organism’s persistence. We could do as-if representational explanation, to some extent, with systems that do not fall into this cluster. But when that worked, it would not be because they exemplify the collection of properties that give content-based explanation its bite. We can still, of course, explain their behaviour in terms of internal workings and how those components are affected by inputs so as to give rise to outputs. As with explaining the trajectory of a falling human in terms of physics and gravitation, that would be to move to a different kind of explanation. It is not because *Homo sapiens* is a historical category that content-based generalizations are inapplicable to the swamp system. It is because the cluster that makes content-related properties project to new cases better than chance is not present. If we do project from real systems to swamp systems, we are relying on currently constituted properties that are much more widely applicable—that apply very liberally in the natural world—and our explanation correspondingly has considerably less explanatory purchase.

Papineau answers the swampman case in a different way (Papineau 2016). He argues that the etiologically based account of content is an a posteriori reduction of our everyday notion. It captures the features which, as a result of scientific discoveries and philosophical theorizing, we have discovered to be important for tying our everyday notion together. If there were lots of swamp creatures around, then our explanatory practices would turn on something else. But in the actual world our explanatory practices are based on the existence of historically constituted properties. Papineau concedes that we could explain and generalize in terms of current properties, and that such generalizations would apply to swamp systems, but he argues that nothing would be gained in the actual world thereby, since swamp systems do not occur here.

My approach is in the spirit of Papineau's observation about an a posteriori reduction. But he shouldn't concede that equally good generalizations in terms of current **(p.171)** properties are available. If there were simply a tie between the explanatory power of currently constituted compared with historically constituted contents, then the absence of swamp creatures in the actual world would not be a decisive consideration. Either pattern would be available as the basis for prediction and explanation. So, I think Papineau's argument needs to be supplemented with the observation that the currently constituted properties that are available for explaining the behaviour of actual creatures and swamp creatures in a unified way are much less satisfactory. They hook on to patterns that exist in those creatures, but that are also found much more widely in nature, and in various degrees. The distinctive explanatory purchase of representational explanation arises because there is a more tightly delineated cluster of properties that arises when consequence etiology gives rise to robust outcome functions supported by internal workings. Swamp creatures fall outside that pattern—if we are restricted to current properties, we cannot explain their behaviour, in this characteristic way, by reference to it.

6.5 Norms of Representation and of Function

(a) Systematic misrepresentation

Another major line of objection to teleosemantic theories is that they do not deliver the kind of normativity that is characteristic of mental content. As I have characterized representational content in the chapters above, the difference between correct and incorrect representations can be captured descriptively. It is a descriptive difference to which norms can readily be applied, just as whether a friend waves or not when we see them is a non-normative fact which can be the basis of praise or censure. Normative properties are not an inherent feature of content. Misrepresentation is one way of explaining a failure to perform task functions. So, if we took it to be a good thing that an organism should fulfil its biological functions, then there would be something wrong with misrepresenting when doing so produces behaviour that fails to fulfil its task functions. But biological well-functioning is just another descriptive distinction. It does not bring genuine normativity into the picture. A descriptive distinction to which norms can be applied is all we should expect in the kinds of cases we have been considering. Norms in a stronger sense—connected to what one ought to do—may arise for representations that are connected to language use or otherwise embedded in a social context, but that is not in play here.

Critics have argued that teleosemantics mistakenly elides misrepresentation with malfunction. That objection needs to be taken seriously even if both distinctions are in the end purely descriptive. Since it is sometimes in an organism's best interests systematically to misrepresent how things are in the world, the objection runs, correctness of a representation cannot be equated with promoting fitness, or indeed with any kind of biological well-functioning. My answer comes in two parts. As we will see shortly, varitel semantics does not equate misrepresenting with malfunctioning. It allows for **(p.172)** malfunctions that are not caused by misrepresentation, for misrepresentations that do not lead to malfunction, and also for misrepresentations that are produced systematically in the organism's evolutionary interests. However, Peacocke puts forward a case which suggests that there is a deeper gulf between misrepresentation and malfunction than my account allows. My answer to that is that we have no reason to think that such cases arise in the kind of subpersonal systems dealt with by varitel semantics. I deal with that first.

Peacocke's example is a case in which a creature systematically misrepresents a predator which is 30 feet away to be only 20 feet away (Peacocke 1993, pp. 224–5). The creature runs away faster as a result and gains a selective advantage by doing so. In that particular example, if there were no other behaviours involved in fixing content, and if the flight response at that speed had indeed been the best trade off of costs and benefits for predators at 30 feet, then Millikan's theory implies that the content is that the predator is 30 feet away (i.e. that it is not misrepresenting). If the details were filled in a bit more so that my account in Chapter 4 applied to this case, then it would have the same result.

These examples do however typically assume that the representation in question is involved in some further pattern of behaviour which fixes its content (see Figure 6.1).[13] That could certainly be the case when we get to human beliefs and desires, which may offer an explanation as to why, in their verbally reported explicit beliefs, human subjects systematically over-estimate the efficacy of their own actions (in contrast with so-called 'depressive realism': the more accurate estimates typically offered by people with clinical depression: Moore and Fresco 2012). If behavioural dispositions to act on a set of representations are formed in one context, and are relatively developmentally fixed, then it may make sense to 'trick' the system when deploying it in other contexts, if the behaviours appropriate to the new context would be different.

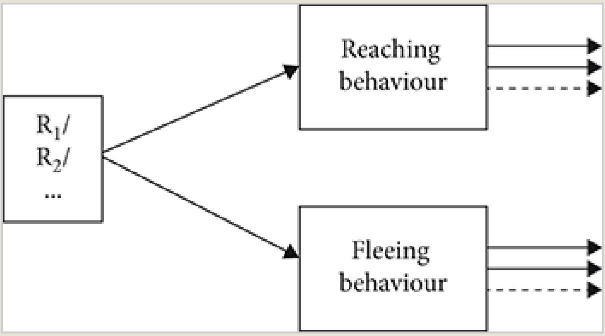

Figure 6.1 The structure of the case from Peacocke (1993, pp. 224–5).

(p.173) However, our simple case studies do not have that kind of structure. In our cases a system has been stabilized for doing a range of things in one context. There is not a second context where benefit is pulling in a different direction so that two different kinds of correctness can become established. Where there are two different routes to behaviour, each of which has been stabilized in different circumstances, then two contents can arise, and they can conflict.[14] Where there is just one route to behaviour, the kind of conflict between correctly representing and well-functioning pointed to by Peacocke does not arise. At least it has not been shown that the intuitive case for the challenge, based as it is on thought experiments where there is more than one route to content, can be extended to our simple cases.

A reason to think that it cannot is that, without further articulation, content ends up being fixed so as to align with whatever story is told about selective or evolutionary benefit. Here is an analogy. Representation theorems in decision theory fix a person's degrees of belief based on the pattern of their choice behaviour. This makes it impossible for a person whose choices obey some basic principles to systematically misrepresent subjective probabilities. A different pattern of choice behaviour would imply a different pattern of degrees of belief. It is only if there is further structure, for example a system of explicit, conscious beliefs expressed verbally by the subject, that those beliefs can come to misrepresent systematically, which in this case would be to say that the explicit beliefs differ from the probabilistic beliefs attributable to the person on the basis of their non-verbal behaviour.

Although I reject the kind of radical disconnect between correctly representing and biological functioning suggested by Peacocke's case, varitel semantics does not elide misrepresentation with malfunction or equate the two. Representational content calls for a constellation of factors, amongst which task functions are just one component. Representing correctly does not reduce to performing task functions successfully, nor does misrepresenting reduce to malfunctioning. (This is the source of the disagreement with Neander discussed in §6.2h above.) A behaviour may have unsuccessful consequences even if the organism is representing everything correctly, for example if something goes wrong with the execution or the environment is uncooperative. Conversely, an organism can misrepresent but through luck produce entirely successful behaviour in accordance with its task functions. Neither distinction reduces to the other. Furthermore, varitel semantics allows for systematic misrepresentation. The costs and benefits of behaviour may be such that selection or learning is tuned to produce many false positives: to be liberal in representing p when p might be the case (Godfrey-Smith 1989, 1991). Then the organism will often be misrepresenting, and acting unnecessarily, although it is operating in just the way evolution has designed it to.

(p.174) In short, varitel semantics allows for some kinds of disconnection between correct representation and biological well-functioning, and there is no reason to think that a more radical disconnection could arise in the kinds of subpersonal systems it targets.

(b) Psychologically proprietary representation

The role of normativity also marks a deep difference between my account and the theory put forward by Tyler Burge in *Origins of Objectivity* (Burge 2010). Burge presents an account of the nature of perceptual representation which, although based firmly in the natural sciences, is very different to the approach I have adopted here. Burge presents his theory in opposition to naturalistic approaches that reduce representational content to a combination of information and function. His first argument against these approaches is that misrepresenting is not necessarily counter to an organism's fitness or biological interests. We have just seen how I address that concern.

Burge's second argument is that teleosemantic accounts are too liberal, setting the border of intentionality too low, and thus allowing in cases where content has no real explanatory value. I disagree. My account of the explanatory purchase of representational content is based on an externalist explanans (one which carries a distinction between behavioural success and failure), together with an explanandum that appeals to externalist properties of internal vehicles (§2.3). It calls for instances of the natural cluster identified in Chapter 3 (§3.2), which is not unduly liberal. But it is true that this account does not depend on the representing system being particularly sophisticated, so it does potentially apply quite widely. This does not set the lower border on representation too low, however. Content-based explanation does indeed have explanatory purchase in this wide class of cases, as I will argue in Chapter 8 (§8.2, §8.5).

Burge makes a third, related argument: that representation proper is proprietary to psychological systems, and is normative. Our accounts of content are thus inadequate for two reasons. They do not capture something distinctively psychological—varitel semantics can in principle apply to non-psychological systems. And they account for content in non-semantic, non-mental, non-normative terms, thereby missing the constitutively normative nature of mental representation.

Given my pluralism, I’m happy to allow that the content of some kinds of mental representation might be importantly different from the subpersonal and from non-psychological cases. Burge may well be right that something different and more sophisticated is needed to characterize the content of personal-level perceptual states; and indeed of beliefs and desires. And there may indeed be something normative going on there, given the way thought and language is embedded in social practices. There is clearly an important difference between our views, however, since I argue that my approach is adequate to account for some genuinely psychological representations, the subpersonal representations in my case studies. I see no good reason to think that subpersonal representational content, of the kind widely relied on in experimental psychology, cognitive neuroscience and the other cognitive sciences, is a kind of **(p.175)** content that is proprietary to the psychological. Indeed, we have good reason to think that it exists more widely, in computational systems that have the same functional profile. Therefore, given our explanatory target, I reject the need for a psychologically proprietary account.

What of normativity? Burge’s approach is deliberately non-reductive: he characterizes what it is to be a representation in terms of having correctness or veridicality conditions, which he characterizes in turn in terms of being a representation, rather than a mere sensory state or informational registration. It is important for Burge that perceptual representations show constancy effects (they are formed by a many-one mapping from sensory inputs). But he doesn’t in the end want to characterize that many-one mapping causally, but to do so in normative terms. A state that is formed in a common way in response to a variety of inputs would not count as showing a constancy unless it had genuine veridicality conditions (i.e. unless it were a representation). Burge is happy for his account to contain this tight explanatory circle because he rejects the need to ‘naturalize’ representational content. He argues that the notion of representation does not need naturalizing, if that requires that representation be explained in other terms. It is an entirely un-mysterious property that plays a central role in the successful science of perceptual psychology and is fully vindicated thereby.

Burge’s thought here would be, I think, a good response to the claim that we should doubt whether there are any representations. Their central role in the sciences of the mind gives us good *prima facie* evidence that there are representations. But my project is directed at a rather different problem: not of showing that there are representations, but of trying to understand their nature better. Burge’s theory does something to characterize their nature, by locating them in a small local holism involving correctness, constancy, and content, but my hope is that we can do more. An account of representational content in terms that are non-mental, non-semantic, and non-normative tells us considerably more about their nature. Of course, it could have turned out that Burge’s account is all that can be said to illuminate the nature of representation. But for the kinds of cases discussed in the foregoing chapters, that would be to give up too soon, since more illuminating accounts of content are available.

6.6 Conclusion

Varitel semantics has several resources for dealing with indeterminacy. It gives rise to more determinate contents than informational semantics or consumer-based teleosemantics. The remaining indeterminacy is a virtue: it is what we should expect in simpler systems with fewer interacting components. Some of the non-conceptual representations in our case studies exhibit some features exemplified by concepts: semantically significant constituent structure; unsaturated components; and limited, domain-specific generality. However, they lack the wide-ranging generality of personal-level concepts. Since the content of a concept is fixed by reference to a wide range of **(p.176)** uses, in combination with many other concepts across many different contexts, conceptual content is likely to be more focused, and thus more determinate, than the contents determined by the simpler interactions and recombinations exemplified in our case studies.

Task functions import a historical component into content determination. That is needed to bring the explanandum into view, the explanandum to which representational explanation is directed, namely explaining successful and unsuccessful behaviour. A history of stabilization operates to constitute some outcomes as being successes—beneficial results—and others as failures, so that even outcomes that are now robustly produced can count as failures in some circumstances. So, we should not expect representational explanation to get a grip, in these simple cases, unless the system has some history of interaction with its environment. However, after even a short period of interaction, some task functions, hence some contents, will begin to be established. So, historically based functions play an ineliminable role in the accounts of content advanced in previous chapters. Varitel semantics does not reduce misrepresentation to malfunction. Misrepresentation does not imply failure to perform a task function, nor the converse.

In short, varitel semantics does a reasonable job of addressing the standard challenges in the literature.

Notes:

[1] Success conditions and satisfaction conditions generally involve the instantiation of a property by a particular. Since we are leaving aside neo-Fregean sense, it does not matter how those properties and particulars are picked out. Different descriptions can pick out the same success condition, e.g. *fly at (x,y,z)* or *fly at my favourite location* pick out the same success condition (assuming I particularly like location (x,y,z) for some reason). Where the property picked out differs, the success conditions are different, even if those properties happen to have the same extension, e.g. *renate animal at (x,y,z) ≠ chordate animal at (x,y,z)*.

[2] I leave aside design function in the following discussion, but appealing to the intentions of a deliberate designer is obviously another way to secure determinacy.

[3] Without computations over internal components, this case has the same structure as animal signalling cases like the honeybee nectar dance (Chapter 1), which are more straightforward for consumer-based teleosemantic views to deal with.

[4] We noted in Chapter 3 that cognitive scientists often take strong correlation to be an indication of what a vehicle represents, but that is because strong correlation is an indication of what a system has evolved or learnt to track. Strength of correlation is not a way of deciding between contents that are equally good from the point of view of explaining stabilization. The evidential test in §4.2c concerns not the strongest correlation, but the correlation changes in whose strength have the greatest impact on achieving task function performance.

[5] It might also fail to qualify as exploitable correlational information, through lack of a univocal reason. Does R carry exploitable correlational information about there being an object b which is S1 or S2 or S3? That requires there to be regions such that P(S1-or-S2-or-S3(b)|R) > P(S1-or-S2-or-S3(b)) for a univocal reason. The property being disjunctive counts against there being a univocal reason. Suppose S1-or-S2-or-S3 forms a proper subset of the category *being a fly* and P(fly(b)|R) > P(fly(b)) for a univocal reason. That same reason is unlikely to connect to the disjunctive category, except because S1 is a kind of fly and S2 is a kind of fly and S3 is a kind of fly. Mentioning this additional factor would make

the reason underlying probability-raising for the disjunctive category non-univocal.

([6]) In many cases that is because a range of mutually incompatible vehicles will carry correlational information about a range of states in the sense defined in §4.1a.

([7]) Cp. the argument in §4.7 that the UE information in the plaid motion system homes in on different correlations for different components.

([8]) Naturalness considerations make that unlikely as between flying insect and flying nutritious object, but the point may apply in other cases, e.g. for locations.

([9]) Note that varitel semantics does not equate misrepresentation with malfunction (see §6.5).

([10]) Martínez (2015) develops and generalizes the view in information-theoretic terms.

([11]) This is related to the 'state view' rather than the 'content view' of non-conceptual representation (Byrne 2005).

([12]) One obvious difference is that the predicative element (the relation of co-activation) cannot be tokened without tokening the singular terms. A natural language predicate (e.g. 'red') can be tokened without tokening a singular term.

([13]) Peacocke mentions another behaviour prompted by these representations: throwing a stone aimed at 20 feet. That would indeed fix a different content, although it is hard to see how the two behavioural dispositions would evolve at the same time. It is more likely that one or other behaviour falls outside the pattern in virtue of which selection has occurred, in which case there is likely to be a directive content which goes unsatisfied, given that the descriptive representation of the situation correctly represents the target predator as being 30 feet away.

([14]) Corollary discharge is in one sense like that (§4.5, §7.4), although there the contents that exist relative to the two different uses are closely related.