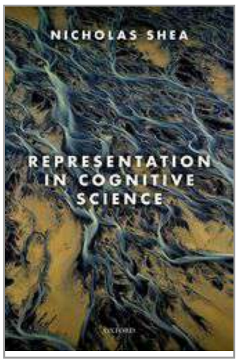

Representation in Cognitive Science

Nicholas Shea

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

(p.227) Paragraph-by-Paragraph Summary

Nicholas Shea

Each entry below summarizes a paragraph of the main text, followed by the page number where that paragraph is found.

Part I

Chapter 1—Introduction

1.1 A Foundational Question

Announcing a mystery: what is thinking, what is a thought process? 3

To Descartes, producing appropriate strings of words was just as mysterious as free will or consciousness. 3

Computers do that. 3–4

Insight: thinking is processing of mental representations. That is RTM. 4

This isn’t an agreed answer to our problem because we don’t know how mental representations get their meaning. 4

There is good evidence for the claim about processing; but I agree we don’t fully understand meaning. 4

Fig. 1.1 Babbage’s difference engine. 5

The question is: how do mental states get their aboutness? 5

That is the ‘content question’. 6

Mental representation is still a powerful idea; there is no serious reason to doubt their existence. 6

But the lack of an answer does lend support to eliminativism, which would be a mistake—it misses something explanatorily important. 6

Fig. 1.2 Typical explanatory scheme from cognitive neuroscience: computation, implemented in brain areas, to perform a task. 7

We leave aside consciousness (too hard). There is non-conscious information processing to get to grips with. 7–8

An answer will also be useful for content disputes in cognitive science. 8

This chapter is a broad overview of existing treatments of the problem (with little argument). 8

1.2 Homing In on the Problem

Brentano identified the problem of intentionality: how can thoughts be about things in the world? 8–9

(p.228) Intentionality of words may derive from our thoughts. The same story cannot apply to our thoughts, on pain of regress. 9

Giving the content of a concept in terms of perceptual states calls for underived intentionality there. 9

We need an account of underived content determination, in the metaphysical sense—a meta-semantic theory. 9–10

I am offering an answer to both the content question and the representational status question at the same time. 10

A theory of content should be consistent with and if possible illuminate behavioural explanation: correct representation explains success, misrepresentation failure. 10

Misrepresentation is puzzling: if it makes no difference to processing, how can it make a difference to explanation? Our theory should show how. 10

The theory needs to be applicable to psychological cases and to deliver determinate contents there. 10–11

The determinacy problem has incarnations as the proximal/distal, qua, and disjunction problems. 11

We are aiming for naturalism: an account in non-semantic, non-mental, and non-normative terms. 11–12

1.3 Existing Approaches

This section discusses existing approaches to content and the main challenges they face. 12

We can appeal to correlations and mathematical information theory (as founded by Shannon). 12

Correlational information on its own is not enough: not determinate; the strongest correlation may not be the right one; and it does not rule out disjunctions. It is just one ingredient. 12–13

Inferential role semantics faces problems of holism, and of identifying sets of inferences that are shared and individuative. 13

Structure mirroring structure is another candidate. Too liberal on its own, but another plausible ingredient. 13–14

Davidson: having beliefs and preferences is a matter of being so-interpretable. 14

Dennett: the intentional stance is applicable because of real patterns, and so is realist. 14

I reserve ‘realism’ for cases where there are real vehicles of content, which is our target, and explains more (see §2.6). 14–15

1.4 Teleosemantics

Teleosemantics is my main precursor. It founds content on etiological functions. Leading theories are also committed to an identifiable representation-consumer playing a content-constituting role. 15

(p.229) Commitment to consumers defines the ‘basic representationalist model’: a consumer conditions its behaviour on an intermediate representation. 15–16

Fig. 1.3 The basic representationalist model. 15

Content of representation R = condition presupposed given the behaviour output by a consumer in response to R. 16

Millikan’s version: content = success condition = condition which explains why the prompted behaviour led systematically to survival and reproduction. 16

This is illustrated by the honeybee nectar dance. 16–17

Fig. 1.4 Honeybee nectar dance. 17

Millikan's 'least detailed Normal explanation' of the pattern of dances homes in on a particular condition for each dance type, e.g. there being nectar 400 metres from the hive. 18

This excludes various general and background conditions from the content. 18

Teleosemantics is the core of a plausible account of content in animal signalling and some other simple cases. 18

1.5 Challenges to Teleosemantics

This section discusses challenges faced by teleosemantics, so my later accounts can be assessed against them. 18–19

First challenge: leading versions require an identifiable consumer whose outputs can play a content-constituting role. 19

It is especially hard to identify consumers in the case of neural representations. 19

Second challenge: to formulate a notion of etiological function that is suited to playing a content-determining role, general enough to cover the cases, and specific enough to do explanatory work. 19–20

Fig. 1.5 Complex functional interconnections in the brain. 20

Relational evolutionary functions do not deliver very specific functions when the function of the learning mechanism is very general (e.g. classical conditioning). 20–1

Instrumental conditioning can found functions in its own right, irrespective of its evolutionary function, as in Dretske (1988). 21

The challenge is to generalize this, and to show the right way to delimit the kinds of etiological functions that should figure in a theory of content. 21

Third challenge: swampman, which raises the question: why should content depend on history? 21

Behaviour is caused by vehicle properties; therefore we could predict the behaviour of an intrinsic duplicate. 21–2

Millikan argues that inductions are based on historical kinds. But that does not explain why we couldn't do induction on synchronic kinds, properties shared between duplicates. 22

Better move: if the explanandum is *successful* behaviour, that is absent in the case of swampman. 22

(p.230) That does not rescue views based on evolution by natural selection. As a swamp creature interacts with its environment, the explanandum will come into place. 22–3

Fourth challenge: how does content get explanatory purchase? See next chapter. 23

Chapter 2—Framework

2.1 Setting Aside Some Harder Cases

Cognitive science relies on representations well beyond the doxastic and the conscious cases. It would be a considerable achievement to give an account of content for these. 25

I set aside four complicating features of everyday representations: consciousness; relations of justification for the person; being reported to others as reasons; and requiring natural-language-like structure. 26

I use ‘subpersonal’ simply as shorthand for representations where these four features are not important to content determination. 26

We now have a wealth of data about representation in the brain against which to test our theories. 26–7

Fig. 2.1 A typical case where neural areas are picked out both anatomically and in terms of what is represented and computed. 27

Neural representations are central cases for us, but the aim is for an account that applies widely in the (subpersonal) cognitive sciences. 27

The overall strategy is to start with the subpersonal, both because this dissolves some of the puzzle of intentionality in its own right, and as a significant step towards a theory of the more complex cases. 27–8

2.2 What Should Constrain Our Theorizing?

We should not rely on intuitions, especially in subpersonal cases. 28

A theory of content should show how correct representation explains successful behaviour and misrepresentation explains unsuccessful behaviour. 28

That requires us to get into the details of a wide range of case studies from cognitive science. 28

Our theorizing is constrained by the need to make representational explanation intelligible. 28–9

Desideratum: an account of content should show why representational properties underpin better explanations of behaviour than would otherwise be available. 29

Content has to be more explanatory than an explanation factorized into a causal chain from input through internal working to output. 29–30

E.g. with a rifle, if the firing pin has content, that explanation marches exactly in step with one in terms of non-representational properties of the causal chain. 30

Demanding that representational explanations be indispensable is too strong. 30–1

(p.231) Fig. 2.2 The mechanism for firing a rifle. 30

The task is to come up with a definition that picks out properties which are realized in the natural world, and to show why the possession of those properties by internal vehicles affords behavioural explanation. 31

2.3 Externalist Explanandum, Externalist Explanans

A commitment to vehicles of content—mental representations as physical particulars—means there will be a full non-semantic causal story. 31–2

For an externalist explanandum—explaining a system’s reaction to and distal effects in its environment—externalist contents can provide an explanation that could not be given in terms of intrinsic properties. 32

The interactions (mappings from distal features to distal outcomes) that call for representational explanation are a subset: performing functions. 32–3

Fig. 2.3 ‘Moniac’—which uses water to compute the way money flows through the UK economy. 33

A function is a mapping. It can be achieved in many ways. 34

An ‘algorithm’ is the way a system achieves a given function. 34

Transitions called for by an algorithm must map onto internal processing going on in the system, non-semantically described. 34–5

E.g. an algorithm to track surfaces by first tracking colour and motion properties of a portion of surface separately. 35

Which extrinsic properties of vehicles determine content? Exploitable relations they bear to the environment. 35–6

Fig. 2.4 An input–output mapping does not fix what goes on inside the box. 35

The algorithmic explanation calls for a convergence between intrinsic properties that account for the processing with extrinsic properties that account for the content. This is the dual nature of content. 36

2.4 Representation Without a Homunculus

It is a mistake to see mental representations as requiring interpretation or an understanding homunculus. 36

Instead, dispositions to produce, process, and act on these vehicles in a certain way, without understanding, is what constitutes them as having content. 36–7

Content arises out of a constellation of interpreter-independent natural properties, so does not depend on the organism being interpretable in a certain way. 37

2.5 What Vehicle Realism Buys

There being a fact of the matter about which algorithm a system is following depends on realism about representation, i.e. upon intrinsically individuable vehicles of content. 37–8

Realism predicts that errors produce further errors downstream. 38

(p.232) Realism explains why representational resources are stable and change piecemeal (when observed). 38

All three patterns of explanation depend on there being non-semantically individuable content-bearers. 38

Vehicles are content-bearers, but vehicle types are not necessarily the content-bearing types; the latter are ‘syntactic types’. 39

Vehicles are individual bearers of content picked out in terms of processing-relevant non-semantic intrinsic properties. Syntactic types are types that are processed the same way by the system, and so are guaranteed to have the same content. 39

Syntactic individuation is partly externalist, but still vehicles within a system share processing-relevant properties, which is what underpins the three patterns of explanation above. 39–40

Dynamical properties can be syntactic properties—interacting in ways that are faithful to their contents. 40

Nothing in dynamicism as such counts against dynamical properties being representational properties. 40

A note on metaphysics of mind. A naturalistic theory of content need not be reductive. 40–1

An illuminating sufficient condition for content can admit of exceptions, exceptions that can only be explained at a more fundamental level. 41

Nor need we get a necessary and sufficient condition for content. A series of different sufficient conditions applicable to different cases would be fine. 41

2.6 Pluralism: Varitel Semantics

My framework has two variable elements: functions and exploitable relations. 41

Exploitable relation is a disjunctive category for me, with no unifying cover-all definition. If a condition that covers both correlation and structural correspondence were too liberal, that would rob representation of its distinctive explanatory purchase. 41–2

Functions are a second source of pluralism. Any theory that covers all four types may well over-generate. 42

Since the definition of task function is disjunctive, it is not really a single condition. 42

A final source of pluralism is that content may well be constituted differently at the personal level. 42

I am not here making a positive argument for pluralism, but resisting the call for a cover-all necessary and sufficient condition. 43

'Varitel' semantics marks the variety of exploitable relations, and of functions, hence the varietal set of accounts of content. 43

Distinctive features of the view: pluralism; restriction to the subpersonal; no consumer; tested not by intuition but another desideratum; and my take on realism, exploitable relations, and the dual nature of content. 43

(p.233) Part II

Chapter 3—Functions for Representation

3.1 Introduction

This chapter focuses on the functions that underpin content, 'task functions'. These are characterized, not by relation to intuition, but by the need to underwrite the explanatory role of representation. 47

Swampman, and the fact that the evolutionary functions of some learning mechanisms are highly unspecific, both motivate a notion of function-for-representation wider than just evolution by natural selection. 47–8

My account combines two elements: producing an outcome in a range of circumstances (robustness) and having a consequence etiology (stabilization). 48

3.2 A Natural Cluster Underpins a Proprietary Explanatory Role

Outcomes being robustly produced and the target of stabilization processes are two features that cluster naturally with a third: having internal representations that produce such outcomes. 48–9

3.3 Robust Outcome Functions

(p.234)

Stabilizing processes can result in a cover-all strategy, where an outcome is produced by just one means, and when it is produced is sensitive to relevant external circumstances, so we should not require that robust outcomes must be produced by a repertoire of different means. 54–5

Sensitivity to inputs does not imply that the organism has to be interacting with an object on which its behaviour is targeted (a requirement that is a weakness of cybernetic accounts of goal-directedness). 55

Pursuing a single means whatever the circumstances does not count. 55

Definition: robust outcome function. 55

'Output' covers bodily movements, actions, and their consequences. 55–6

What counts as a different input for clause (i), and not just a different instantiation of the same type of input, is a subtle issue. 56

The idea of 'relevant' external conditions in clause (ii) also needs careful handling. 56

3.4 Stabilized Functions: Three Types

(a) Consequence etiology in general, and natural selection

Stabilized functions are based on natural selection and/or learning and/or contribution to an organism's persistence. 56

'Consequence etiology' is where an output F is generated because of its consequences: F occurs because it is produced by a system S, and S is present because it produces or has produced F. 56–7

Fig. 3.3 General treatment of consequence etiology. 57

This draws the category of function too broadly for our purposes. 57

I draw it disjunctively. Evolution by natural selection is the first case. 57

(b) Persistence of organisms

A ubiquitous enabling condition for producing an outcome F robustly is the persistence of an organism disposed to produce F. 57–8

Several accounts of biological function base it in contribution to a system's persistence. 58

To define the class of systems, contributions to whose persistence should count, I help myself to the category *organism*. 58

(p.235) Bacterial chemotaxis exemplifies contribution to the persistence of an organism. 58–9

An outcome F which is like that occurs partly because of the consequences which flowed from the individual organism producing F in the past. 59

The effect is not specific to F (unlike natural selection or learning), since it preserves S with all its behavioural dispositions. Persistence is an indirect route, when combined with learning, towards making F more robust. 59

(c) Learning with feedback

Learning need not be driven by outcomes that contribute to persistence of the organism. 59

With feedback-based learning, the consequences of producing F (e.g. of pressing a key) account for the disposition to produce F (in certain circumstances), and to do so robustly. 59–60

Fig. 3.4 Subjects learn from feedback to make reaching movements that maximize reward. 60

Learning is a form of diachronic flexibility in the service of being able to achieve important outcomes robustly. 60

Learnt behaviours have derived evolutionary functions, derived from the function of the learning mechanism, e.g. to track people by their faces. 60–1

Derived evolutionary functions can be quite unspecific, e.g. when learnt by classical conditioning. 61

A learning process can explain stabilization without explaining why certain outcomes (e.g. getting money) reinforce/modulate dispositions to behaviour. 61

This encompasses one-shot learning, negative reinforcement, and cases where achieving outcomes close to O increases the organism's tendency to achieve O. 61–2

Behaviour learnt by imitation may involve feedback, or may have a stabilized function as a result of cultural transmission and selection. 62

In sum: feedback-based learning is a third source of stabilized functions. 62

(d) A 'very modern history' theory of function

Stabilizing processes are not like kinematic equilibria: forces holding a pattern in place at a time. 62So, it is tempting to treat them counterfactually or in terms of what is likely to be stabilizing. 62

Whether an outcome will contribute to persistence (or learning or natural selection) is unsuitably open-ended. 62–3

Forward-directed functions are also poorly suited to being part of an explanation of why outcomes are produced; historically based functions can. 63

So, I define stabilized function in terms of an actual causal history of natural selection, learning or contribution to an organism's persistence: a 'very modern history' theory of function. 63

Definition: stabilized function. 64

Evolution covers cases of cultural transmission and selection. 64

(p.236) Outputs on the way to, or downstream of, some output F that is stabilized do not thereby also count as stabilized functions. Outcomes that have contributed to stabilization unsystematically also don't count. 64

3.5 Task Functions

The functions-for-representation that will figure in our theory are task functions. A task function is a robust outcome function that is also a stabilized function or the product of intentional design. 64

Functions based on intentional design do not meet our criteria for naturalism, so while noting that they share features with the others, I mostly set them aside. 65

I also set aside representational contents that are the direct result of design, as when a designer intends that x should represent y. 65

Definition: task function. 65

This is not offered as a definition of biological function, and I don't follow previous theorists in aiming to reduce the supposed normativity of content to the supposed normativity of biological function. 65

Task function works in the subpersonal cases we will examine. Other notions of function (or none) may be needed elsewhere. 65–6

Task functions vary in ways that affect the explanatory bite of the contents they generate. For example, robustness and stabilization come in degrees. 66

They also vary in how many of the alternative bases of stabilized function are present: all three in paradigm cases, but they need not line up and may underpin different (and contradictory) contents. 66

In the balance between generality and informativeness, our cluster is encountered often in nature while still having the benefit of underpinning a rich set of inductions. 66

3.6 How Task Functions Get Explanatory Purchase

(a) Illustrated with a toy system

This section looks at a toy system based on the comparator circuit for motor control of reaching. 67

Our system will move along a line, from a range of initial positions, to a set position *T*, where it stops. 67–8

Fig. 3.5 The toy system. 67

We can explain how the system achieves that outcome by reference to the processing of internal components that bear certain exploitable relations to features of its environment. 68

Reaching *T* is a robust outcome function. 68

If it recharges at *T*, then reaching *T* also becomes a stabilized function; also if it learns, so that recharging cements an initially randomly set behavioural disposition. Reaching *T* is then a task function. 68

(p.237) An interaction between four internal components, given how they correlate with things in the environment, explains how the system achieves this task function. 68–9

Successfully reaching *T* is explained in terms of components correlating as they did during stabilization, i.e. representing correctly; failure to reach *T* is explained in terms of misrepresenting. 69

(b) Swamp systems

What if our toy system were assembled by chance? It would have a robust outcome function, e.g. reaching the location *T*, but none of the possible robust outcomes seems to count as a success. 69

Suppose there happens to be a power source at *T* and the system has been moving around for a while, during which time it has recharged at *T*. Reaching *T* again would then count as a success. 69

Going beyond that intuition, we can see that the success/failure distinction arises because of our cluster. Then the successes are outcomes for which it is explicable both why and how they are produced. 69–70

The same goes for organisms: a robust outcome function of a swamp monkey would not count as success or failure until the individual’s interactions with the world had contributed to persistence or learning. 70

These thought experiments illustrate how task functions, and the success/failure distinction, can arise independently of evolutionary history. 70

Learning-based task functions are the same, e.g. clapping because it makes a parent smile. 71

Task functions still depend on history, but unlike standard teleosemantics, not on deep evolutionary history. 71

The robust outcome aspect of task functions means that there are real patterns in system–world interactions that generalize across distinct proximal properties. That contributes to the proprietary explanatory purchase of representational content, cp. §8.2b. 71

Unlike in the case of the rifle firing pin (§2.2), robust outcome functions ‘bridge’ to common outcomes across a range of proximal conditions. 71

Standard teleosemantic accounts do not include a robust outcome requirement, although their motivating examples would support it. 71–2

3.7 Rival Accounts

Griffiths argues that function should not be analysed in terms of contribution to persistence, because natural selection is aimed at reproduction and that can work against persistence of the individual. 72

Our account still applies in such cases. Task functions will be based on evolution rather than persistence. 72

(p.238) Griffiths offers a rival forward-looking account. 72–3

The objections above to forward-looking accounts (§3.4d) apply equally here: they are too open-ended, and are unsuited to figure in a causal explanation of behaviour. 73

A consequence of my account is that representations based on an evolutionary history of stabilization (reproductive fitness) can conflict with those based on persistence of the individual (survival).

Task functions can also arise based on contribution to persistence without any evolutionary benefit. 73–4

3.8 Conclusion

Three features cluster together for a natural reason: robustness, stabilization, and processing over internal components bearing exploitable relations. This constitutes some behavioural outcomes as successes and allows us to explain both how and why they are produced. It is this cluster that gives rise to the kind of representational content found in our case studies. It allows us to see why content has a distinctive explanatory role, thus satisfying our desideratum (§2.2). 74

Chapter 4—Correlational Information

4.1 Introduction

(a) Exploitable correlational information

The next two chapters turn to the internal algorithm; in this chapter, where correlation is the relevant exploitable relation. 75–6

The account does not call for dedicated representation consumers that play a content-constituting role. 76

Definition: correlational information. 76

Correlations are useful. Behaviour can be conditioned on a correlate. 76

The exploitable correlations—ones that can be relied on, and where things going well is explicable rather than accidental—are nomologically underpinned by a single reason. 76–7

Definition: exploitable correlational information. 77

The regions in which a correlation subsists can be very local. 77–8

Exploitable correlational information may be carried by a range of states about a range of states. 78

Definition: exploitable correlational information carried by a range of states. 78

Cases of exploiting a correlation: animal signalling; equilibria in Skyrms-style signalling games. 79

Exploitable correlational information is liberal: it exists with respect to many different regions, with different strengths. 79

The relevant correlation strength is that within the region encountered by organism/system (individual or type). 80

(p.239) (b) Toy example

We will look at a toy example to see the way correlations are exploited. 80

Table 4.1 Useful correlational information carried by the four internal components. 81

Fig. 4.1 Toy system discussed in the text. 81

Given the correlational information carried by the internal components, the way they are processed constitutes a simple algorithm for moving to a power source.

They meet our desideratum for the explanatory role of content. 81

The components carry other information, e.g. about sensory stimulation or rotation speed of the wheels, but that information cannot explain task performance so directly. 82

Correlation of internal components with light intensity would also offer a less direct explanation. 82

Some content indeterminacy remains, e.g. as between distance to a power source and distance to something worth reaching. 82

4.2 Unmediated Explanatory Information

(a) Explaining task functions

Our motivation for representationalism implies that the content-constituting correlations are those which explain S’s performance of task functions. 83

We are not asking which content assignment gives the best representational explanation of behaviour. We are asking which correlations offer the best causal explanation of robustness and stabilization. 83

To explain S’s performance of task functions is to explain how outcomes were stabilized and robustly produced. 83–4

Definition: explanandum: S’s performance of task functions. 84

Definition: unmediated explanatory information (UE information). 84

A correlation with C plays an unmediated role in the explanation if its role does not depend on C correlating with some further condition C’. 84

UE information constitutes content. 84

Condition: for content based on correlational information. 85

The theory closely parallels the use of model-based fMRI to find representations in the brain. 85

To choose between candidate algorithms that would account for the observed behaviour, we examine which algorithm best accounts for trial-by-trial variations in neural activity. 85

As well as keying into external conditions, explanatory correlational information also has to fit with internal processing: with (locally caused) transitions between the information-carrying vehicles. 85–6

Algorithms usually call for processing steps in which different vehicles have different contents. ‘Unmediated’ does not count against the computation being mediated by a series of steps. 86

(p.240) UE information covers output correlations too. These give rise to directive contents. But there has to be some descriptive content in the system somewhere. 86

There are some internal states whose tokening correlates with achieving an output that is itself a task function, e.g. getting sugar. When it forms part of a wider story, that will count as UE information. 86–7

My account is in the spirit of Dretske (1986, 1988), where correlational information is a structuring cause of behaviour. 87

But my account is more general: it is not restricted to instrumental conditioning; correlations need not pre-exist; and it applies when several vehicles need to interact to produce the stabilized behaviour. 87–8

The latter is important because real cases have many representational vehicles that interact in complex ways. 88

(b) Reliance on explanation

As in other sciences, I am assuming a realist account of causal-explanatory relations. 88

So, it is not an interest-relative matter whether UE information exists. It may be interest-relative whether we go in for representational explanation (i.e. appeal to the property we have theorized). 88

If causal-explanatory relations are interest-relative throughout the sciences, then they would be in my theory too. 88–9

We do need to ask whether the property picked out by the definition of UE information is any use. It is, because it is in fact the property that figures in many representational explanations of behaviour. 89

(c) Evidential test

UE information makes performing task functions more likely. That gives us a test for UE information. 89

Evidential test for UE information. 89

In the toy system, strengthening the correlation of **r** with location directly affects the chance of reaching *T*. 89

Strengthening the correlation with light intensity would have a less strong effect on achieving that outcome. 89–90

The test also applies to correlations at output (e.g. noise in the motor system). 90

It is only an evidential test, not always satisfied, and not necessarily giving the right result. 90

The test is restricted to correlations with conditions involving natural properties, since these are the best candidates to figure in causal explanations of a system's performance of task functions. 90

To apply the test, keep fixed how the vehicles interact, and consider what would happen if the world was in a particular state more or less often. 90–1

(p.241) The test can be applied to the system in the past (during stabilization) or, less ideally, applied to current circumstances. 91

4.3 Feedforward Hierarchical Processing

Here we look at a simple case: the ALCOVE feedforward neural network. 91

Its task function is to put objects of category A into box A. 91

Nodes at the hidden layer ('exemplar nodes') correlate with individual objects and carry information about many other things (i)–(vi). 91–2

Fig. 4.2 The ALCOVE network. 92

At the hidden layer, the unmediated explanatory information is about exemplar identity. 92

According to the evidential test, the UE information at output concerns object categories; at input, perceptual features. 93

At the hidden layer, exemplar and category are tied on the evidential test, but exemplar content gives us a better understanding of how the system achieves its task functions. 93

UE information chooses between coextensive correlations and tends to home in on distal properties, e.g. in another model JIM, the geons are about objects rather than sensory features. 93

A further development of ALCOVE has reciprocal connections. We turn to feedback in §4.8. 93

This account of content makes good on my earlier claim that there is no need for a representation consumer that plays a content-constituting role. 93

4.4 Taxonomy of Cases

Neural processing takes place in complex ways. 94

Fig. 4.3 Diagram of the primate visual system. 94

Fig. 4.4 The four kinds of case exemplified in §§ 4.5, 4.6, 4.7, and 4.8 respectively. 95

I will pick out four kinds of cases and look at an example of each. 95

It is very hard to identify a single consumer system when there are feedback loops and no simple hierarchy. 95

In Case 1, one vehicle is used by two different output subsystems. In Case 2, two different representations are used by a single subsystem. 95–6

In Case 3, information is processed via two routes, one direct and one indirect. Case 4 features feedback and cycles. 96

The case studies below consider each in isolation to show that none presents an obstacle to the UE approach. 96

4.5 One Vehicle for Two Purposes

It is common to find one animal signal used by different receivers for different information it carries (e.g. mates and predators of a firefly). 96

(p.242) A cooperative example of that is a chicken signalling to conspecifics and to a predator. 96

Corollary discharge tells the motor system to act and tells perceptual systems that the organism is acting. 96–7

C. elegans contains a simple example of this. 97

Arguably the contents are of different kinds here, descriptive and directive (see Chapter 7). 97

In our case studies, where there is reuse it turns out that the same content is being used by different subsystems. 97

4.6 Representations Processed Differently in Different Contexts

(a) Analogue magnitude representations

The analogue magnitude system looks at first as if it represents different contents in different contexts, but actually there is probably a common representation of numerosity. 97–8

It acts as an (imperfect) correlate of the numerosity of many kinds of array, with discrimination displaying the characteristic Weber's law set-size signature. 98

There is evidence for a common numerosity register—comparisons across modalities, interference, common neural basis—while also registering whether the items are objects, tones, flashes, etc. 98

Consider a stylized system with one register R for numerosity and another R’ for stimulus type. 98–9

Fig. 4.5 Case 2. 99

Rather than separate domain-specific contents for different downstream uses, the UE approach implies that R is a common register for numerosity *tout court*. 99

These considerations mean that UE information will often ‘triangulate’ on a common content. Being decoupled from specific outputs, perceptual representations are pushed towards having purely descriptive content. 99–100

The systematic relationship between activation and numerosity means that the system can represent numerosities beyond those encountered during stabilization. 100

(b) PFC representations of choice influenced by colour and motion

We turn to the prefrontal cortex for a case where the two different kinds of information are carried in the same register and used in different contexts. 100

Subjects see an array of random dots in varying proportions of two different colours and with varying motion directions. Their task is either to judge preponderant colour or preponderant motion direction. 100

Fig. 4.6 Behavioural task in Mante et al. (2013). 101

Neural evidence accumulates for both colour and motion, in a distributed neural pattern. 101

(p.243) The context affects whether the evidence that accumulates to drive choice (movement in the choice direction) is colour or motion. 101–2

Fig. 4.7 Schematic representation of the processing in Mante et al. (2013). 102

Simplify: representational processing of two vehicles at input, one at output (programming saccades). 102

Task function: to get juice; relevant condition is colour in some trials, motion in others. 102

List of correlations with distal features that explain performance of the task function, hence constitute content. 102

Correlations with properties of sensory input would be less explanatory. 102–3

Lumping all the inputs together in a single space is less explanatory of how the system computes. 103

Some indeterminacy remains. That is appropriate to the case. 103

4.7 One Representation Processed via Two Routes

Case 3 is where a single representation is processed by two routes, one direct and one indirect. 103

We look at an example of that in the visual system, from van Essen and Gallant (1994). 103–4

Fig. 4.8 Case 3. 103

Fig. 4.9 A detail from van Essen and Gallant (1994). 104

We focus on the interaction between V2 Thin stripe, V2 Thick stripe, and MT, with a simplified computational interpretation. 104

MT detects plaid motion: motion of overlapping surfaces moving in different directions. 104

There are two routes from V2 Thin stripe to MT. 104–5

We consider a simplified system with the task function of intercepting moving objects. 105

MT correlates with direction of motion of encountered objects, V2 Thick stripe with local motion direction, and V2 Thin stripe with a chromatic invariant of surfaces. 105

Different components are doing different jobs. 105

A consumer-based approach could bundle these together into a single intermediate representation, but that would not account for how the system manages to compute motion direction. 105–6

In short, the UE information approach is appropriate for cases of one vehicle processed via two routes. 106

4.8 Feedback and Cycles

Bogacz (2015) describes a probabilistic calculation for deciding between options on the basis of sensory evidence. The model involves cyclical information processing en route to action selection. 106

(p.244) The circuit calculates probabilities that various possible actions will be rewarded. When one of these crosses a threshold the system outputs the corresponding action. 106

To see how the UE approach can apply to probabilistic representations, first notice that fine-grained information about the probabilities of a range of world states can be useful information to have. 106–7

Fig. 4.10 The neural computation proposed by Bogacz (2015). 107

The joint probability distribution of a range of putative representations X and a range of world states Y, if based on a univocal reason, is a *fine-grained exploitable correlation carried by X about Y*. 107

Correlations like that, which figure in explaining performance of task functions, qualify as UE information, hence fix content. 108

Conditions Y about which X carries more mutual information are, *ceteris paribus*, better candidates to be UE information carried by X. 108

We need to replace correctness with a graded notion of accuracy. The Kullback-Leibler divergence of the true distribution from the represented distribution is one appropriate measure. 108–9

In Bogacz (2015), the system takes a sensory input and uses it to calculate posterior probabilities that available actions will be rewarded. If none crosses a threshold, they act as priors that are updated with the next sensory input. 109

The system has been trained by feedback to produce the action that is mostly likely to be rewarded in the current context. 109

Bogacz's probabilistic computational model captures the UE information used by the system to perform the task. 109–10

Fig. 4.11 Case 4. 109

This case shows that the varitel framework can discern content processed in feedback loops. 110

4.9 Conclusion

For a relevant input–output mapping (and Chapter 3 told us that task functions are the relevant ones), content is fixed by exploitable relations carried by components which make internal processing an implementation of an algorithm to perform the input–output mapping. 110

For correlations, the content-constituting ones are those which unmediatedly explain, through implementing an algorithm, how a system performs its task functions. That works in a series of case studies. It does not call for a content-constituting consumer system. 110

Chapter 5—Structural Correspondence

5.1 Introduction

This chapter: structural correspondence is an exploitable relation and can be content-constituting. 111

(p.245) A correspondence: maps entities from domain 1 to entities in domain 2; and there is a relation between the entities in domain 1 that is mirrored by a relation between the entities in domain 2 to which they map. 111–12

The thin notion of relation makes the existence of a structural correspondence between relations too undemanding. Here I develop principled restrictions to candidate relations on both sides of the correspondence. 112

For every relation on entities in the world, and any way of mapping a set of vehicles onto them, there is a corresponding relation on the vehicles. 112

In general, a structural correspondence of this kind is not something that will help a system perform task functions. So, it cannot be content-constituting while meeting our desideratum. 112

Fig. 5.1 For any relation on entities in the world, there is *a* corresponding relation on vehicles. 113

The cognitive map in the rat hippocampus exemplifies a more substantive kind of correspondence, an *exploitable structural correspondence*. These can be exploited, thus constituting content, or go unexploited. 113

5.2 The Cognitive Map in the Rat Hippocampus

Rat place cells fire when the rat is at a specific location. 113–14

Fig. 5.2 Place cells in the rat hippocampus are tuned to specific locations. 114

This is useful information to have, e.g. to learn what to do in different locations, but that would not be to make use of a relation between place cells. 114–15

Taken on its own, place-specific firing does not show that any relation on the place cells is being exploited. 115

Cells for nearby locations tend to activate one another when active offline, showing replay or preplay of routes through space. 115

Co-activation is used to compare different routes to a rewarded location and pick the shorter one. 115–16

This is a matter of achieving task functions robustly, stabilized by interaction with the environment, part of the explanation for which is the structural correspondence between co-activation and space. 116

So, this is a case where use of a structural correspondence to perform task functions is the basis of representational content. 116

5.3 Preliminary Definitions

This section defines 'structural correspondence' and 'structural representation', and says what it is for a structural correspondence to be content-constituting. 116

Symbols: the structural correspondence obtains between relation V on vehicles v_i and relation H on worldly entities x_i. 116–17

Fig. 5.1 shows an isomorphism, but I define structural correspondence in terms of the slightly looser notion of homomorphism, which allows for two different representations with the same content. 117

Definition: structural correspondence. 117

(p.246) This does not imply that the parts have to be representations. For simplicity I use the standard definition of structural representation, which assumes that they are, but my approach could still apply if they're not. 117–18

What it takes to be a structural representation is that a relation on representations represents a relation on the entities represented. 118

Definition: structural representation. 118

We are interested in cases where a relation on vehicles represents a relation in the world *because* the relation on vehicles bears a structural correspondence to the world. The definition of structural representation does not entail that. 118

Definition: structural correspondence as content-constituting. 118

For a system to make use of a structural correspondence, the relation V between vehicles has to make a difference to downstream processing. 119

For contrast, consider vervet alarm calls and, somewhat arbitrarily, the relation H, *higher than*, between predators (i.e. how high off the ground the predator usually is). 119

A relation on the alarm calls corresponds to H, but vervets are not sensitive to that relation (nor to any relation between alarm calls). The structural correspondence exists, but is not content-constituting. 119

The requirement that a structural correspondence be used, therefore usable, cuts down very considerably on the problematic liberality of structural correspondence. 119

5.4 Content-Constituting Structural Correspondence

(a) Exploitable structural correspondence

This section sets out the substantive notion of structural correspondence we need. 120

In the rat case, co-activation made a difference to processing, and its correspondence to distance was being used. 120

Having a relation that processing is sensitive to correspond to a task-relevant relation in the world is a major achievement. 120

Definition: exploitable structural correspondence. 120

Downstream processing is sensitive to co-activation, but not to the colour of the cell bodies, nor to where they are located within a hippocampal layer. 120–1

The place cells on their own are useful because they enable acquisition of a co-activation structure, but I reserve ‘exploitable structural correspondence’ for when the relation on vehicles is already in place. 121

NB: the exploitable relation is not the co-activation relation. It’s the overall structural correspondence. 121

Neural processing is sensitive to relations between firing rates, and temporal relations between spikes; possibly also to phase offset. 121

Plasticity can alter downstream processing so that a merely potentially exploitable structural correspondence turns into an exploitable one, i.e. one where processing is systematically sensitive to the relation between vehicles. 121–2

(p.247) The relation on vehicles should make a systematic difference to downstream processing—which can be spelt out. 122

The worldly relation has to be significant for the system, given its task functions. That will usually exclude gruesome and disjunctive properties. 122

Notice that there are different restrictions on the two sides of the correspondence. 122–3

A structural correspondence is instantiated when an instance of the relation V is instantiated between two vehicles, together with an instance of relation H being instantiated between the two worldly entities to which they correspond. 123

We have identified the useful structural correspondences. This captures the sense in which the Survey of India was useful. 123

(b) Unmediated explanatory structural correspondence

For an exploitable structural correspondence to be exploited is for it to figure in a causal explanation of a system's performance of task functions. 123

Definition: unmediated explanatory structural correspondence (UE structural corres- pondence). 123

For the rat, getting to a particular location again is a task function, performed by using the structural correspondence between co-activation and space. 124

This gives rise to content based on UE information and UE structural correspondence. 124

Definition: condition for content based on structural correspondence. 124

So structural correspondence, of an appropriate kind, is part of what gives representations their content. 124

The definition is neutral between descriptive and directive content—see Chapter 7. 124

Exploitable structural correspondence is not (circularly) defined in terms of being exploitable. 125

A UE structural correspondence can determine content about entities x_n and a relation H on them all at once. 125

Fig. 5.3 Points on a simple map pick out locations, and do so in virtue of their spatial relations to other things on the map. 125

A new exploitable structural correspondence can be created by creating new relations amongst existing vehicles, e.g. learning the (arbitrary) count words by rote. 126

Putting together a series of vehicles so that the sequence becomes automatized is a common way to create new structure. 126

In short, new exploitable structural correspondences can be created by constructing new relations on vehicles or by making downstream processing newly sensitive to an existing relation on vehicles. 126

5.5 Unexploited Structural Correspondence

In some cases an obvious structural correspondence is not being exploited and is not a basis for content. 126–7

(p.248) Many deliberately designed representations are set up so as to make an obvious relation usable, e.g. spatial relations. 127

Colour is another easily used relation on representational vehicles. 127

Many cognitive science cases are like the honeybee nectar dance in that a structural correspondence which obviously exists is not being exploited. 127

The bee dance is, however, an ‘organized sign system’ (Godfrey-Smith 2017). 127–8

Organization in this sense is an important feature. It allows a compact mechanism to extend to a range of cases (cp. §4.1a), and to extend to new cases. It also makes the system error-tolerant. 128

Organization is different from structural representation. 128

Cummins’s ingenious driverless car example is a case of structural representation. The structure is the relation between subsequent pin positions on the guide-card. 128

Fig. 5.4 Cummins’s case of a driverless car guided by a slot in a card. 129

The distance moved by the pin correlates with how far the car has moved on the ground. The car uses spatial relations between positions on the card to program appropriate actions. 129

Fig. 5.5 One step of the computation being performed in Cummins’s driverless car case. 130

This is a case of UE structural correspondence: relations between pin positions on the card represent relations between locations of the car on the ground. 130

Gallistel’s concept of a functioning isomorphism inspired, and is much like, my notion of UE structural correspondence, but is more permissive in one important respect. 130–1

Gallistel allows that ‘indirect’ isomorphisms, ‘created only by way of an interpretative code’, are a sufficient basis for content. 131

That is too liberal. The interpretative code could work on each representation piecemeal. Then relations between representations would lose any significance for content. 131

I agree that which relations count depends on downstream processing, but this must be a matter of sensitivity in a systematic way to some interpreter-independent relation on the representations. 131

So, I allow some but not all of Gallistel’s indirect isomorphisms. 131–2

5.6 Two More Cases of UE Structural Correspondence

(a) Similarity structure

We examine two more case studies in which a structural correspondence is exploited and is thereby constitutive of content. The first is exploitation of similarity structure. 132

Similarity in neural activation patterns can be measured by distance in activation state space. 132–3

(p.249) Fig. 5.6 Illustration of neural similarity space. 132

A neural similarity structure, e.g. as observed in BOLD repetition suppression effects, is relevant if subjects' task is to judge the similarity of objects presented to them. 133

When similarity in neural activation space is being used for the way it corresponds to some objective dimension(s) of similarity between objects, this is a case of UE structural correspondence. 133–4

Note: this is not based on a claim about subjectively experienced similarity. 134

(b) Causal structure

The second case involves representation of causal structure, which has been so significant for human evolution. 134

Action selection can be based simply on whether an action led to reward in the past (model-free) or on an understanding of the causal connections between actions and their consequences (model-based). 134–5

The two-step task tests for model-based choice, hence causal understanding, but does not require structural representation. 135

Huys et al. (2012, 2015) tested causal planning using a more complex multi-step task. 135

This causal planning capacity may be an elaboration of the capacity to represent the sequential order of events. When the sequential structure mirrors causal structure, that correspondence is exploitable to do causal reasoning. 135–6

Fig. 5.7 The structure of the task studied by Huys et al. (2012, 2015). 136

If subjects doing causal planning rely on sequential order between brain states corresponding to causal accessibility between world states, then this is a case of UE structural correspondence. 136–7

5.7 Some Further Issues

(a) Exploiting structural correspondence cannot be assimilated to exploiting correlations

Often the representing relation will also correlate with what it represents. Is correlation doing all the content-constituting work? 137

Using a relation between vehicles ≠ using it because it corresponds to a relation between the entities represented by those vehicles. Example: the difference between the firing rates of neurons representing the gaze direction of each eye correlates (inversely) with the distance of the object of focal attention. 137

Fig. 5.8 The difference in firing rate between gaze-direction neurons for each eye correlates inversely with the distance to the attended object. 138

But this is being used for its correlation with object distance, not because it corresponds to a relation between the things represented by the two gaze-direction neurons. 138

(p.250) A second objection is that the cases of computations involving UE information in the last chapter already depend on structural correspondence: that the functional transitions in the computation correspond to the structure of the world. But that is not a case of structural representation at all. 138–9

UE structural correspondence is a special kind of case, with two consequences:

(i) add a new vehicle to the structure and it acquires content irrespective of any correlations;
(ii) the relation can be used to compute across a range of vehicles in a systematic way. 139

An entirely accidentally corresponding structure could be used for its correspondence to the world. 139–40

In sum, UE structural correspondence is a separate basis of content from UE information. 140

(b) Approximate instantiation

Correspondences that are only approximately instantiated can still explain performance of task functions. 140

A structural correspondence I is *approximately instantiated* when the actual relation between the objects represented is approximately equal to their relation under I. 140

Since we have put strong constraints on the objects and relation that can figure in I, it is very unlikely that I is ever exactly instantiated. 140

Allowing approximate instantiation opens up a wide class of candidate exploitable structural correspondences. We need to ask how exactly/approximately each was instantiated. 140–1

The degree of approximateness should match the degree to which behaviour contributed to stabilization. 141

This does not trump considerations about which objects and properties are explanatory of task performance. 141–2

Representational redundancy, where two vehicles map to the same object, is possible in a candidate structural correspondence, but this will reduce the accuracy with which the relation between objects in the world is represented. 142

(c) Evidential test for UE structural correspondence

With this notion of approximate instantiation we can formulate an evidential test for content (cp. §4.2): the correspondence whose accuracy of instantiation is most directly connected to the likelihood of achieving task functions is a good candidate to be the content. 142

Evidential test for UE structural correspondence. 142

The test is epistemically helpful in dealing with indeterminacy. 143

The test does not imply that the more accurate correspondence is always the better candidate. 143

(p.251) Applied to the similarity space on bird images in Constantinescu et al. (2016), the evidential test delivers the content assignment argued for above. 143

5.8 Conclusion

The liberality of isomorphism is a problem because it would rob representation of its explanatory power. From our perspective, liberality is a symptom of a deeper problem: most correspondences are not usable, let alone used. This chapter delimited a restricted class of exploitable structural correspondences. We saw that these are a plausible basis for content determination. 143–4

Part III

Chapter 6—Standard Objections

6.1 Introduction

The aim of this chapter is to make more explicit the way varitel semantics deals with standard philosophical challenges faced by theories of content. 147–8

6.2 Indeterminacy

(a) Aspects of the problem

The plan is to look at determinacy problems in the standard case (the frog prey-capture system), the analogue magnitude system and, in subsection (d), cognitive maps in the rat. 148

We'll work with a simplified version of the frog's tongue-dart prey-capture system. 148

Consider representation R, in the retinal ganglion, which responds to flies at location (x,y,z) and causes a tongue dart to that location. 148

Distality problem: does content concern the fly and the action of catching it, or more proximal stimulation and bodily movement? 148

Specificity problem: which of a nested set of co-instantiated properties is represented? 149

Disjunction problem: why is content not a disjunction of relevant conditions? In my terminology, all three are aspects of the determinacy problem. 149

Answers are tested, not against intuitions about the case, but by whether they deliver the right amount of determinacy for explaining content-based explanation of behaviour. 149

(b) Determinacy of task functions

First move: task functions contribute some determinacy. 150

Task function and a (simple) algorithm are in place in the case of the frog. 150

(p.252) Its task function is to catch flies, not little black moving things—based on which properties are directly responsible for survival and reproduction (selection for). 150

This also counts against background conditions (e.g. strength of gravity) and fine-grained qualifications (e.g. not poisonous) figuring in the content. 150–1

The task function is however indeterminate as between fly (biological taxon), flying nutritious object (ecological category) and object worth eating. 151

In the analogue magnitude system, its (learning-based) task functions home in on *numerosity*. 151

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

The requirement for unmediated explanation of task functions also contributes some determinacy. As a result, distal conditions are often better candidates. 151–2

In the analogue magnitude system, correlation with numerosity explains directly. Correlation with sensory properties figures only in a mediated explanation. 152

Asking how a collection of correlations explains an ensemble of task functions also helps, e.g. it homes in on flies rather than food in general. 152

Varitel semantics cannot appeal, for determinacy, to facts about what the organism can discriminate; nor to which correlation is strongest. 152

Convergence between correlational information and task function provides some constraint: exploitable correlational information must be nomologically underpinned, for a univocal reason. That favours a non-disjunctive category like *fly* over a disjunctive category like *species 1 or species 2 or species 3*. 153

Indeterminacy remains, e.g. as between fly and flying nutritious object, and as between various ways of precisifying the category *fly*. 153

Driving towards determinacy in the analogue magnitude case is the fact that the representations are used across a range of different downstream computations and behavioural outputs. 153–4

The account does not require a representation to be caused by what it represents. 154

Pietroski's case does not prove the contrary once we consider simple non-conscious systems and foreswear problematic intuitions. 154

(d) UE structural correspondence

In the rat, the relevant correspondence is with distal features, for the reason we have just seen, but there may be indeterminacy between various ways of understanding what locations are, e.g. as between absolute and relative spatial positions. 154–5

It may be indeterminate whether place cells pick out locations indexically or non-indexically. 155

(e) Natural properties

The approach tends to favour natural properties. That counts against arbitrary disjunctions. 155

(p.253) Such properties tend not to figure in causal explanations, which stops Peacocke-style reduced content arising in these simple systems. 155–6

More complex representations, like human conceptual representations, clearly can represent such contents, using the combinatorial power of concepts. 156

(f) Different contents for different vehicles

A soft constraint deriving from the definition of UE information is that different vehicles in the same range should generally have different contents. 156

E.g. with the frog, ascribing the content *there is a fly nearby* to the firing of all retinal ganglion cells would be less explanatory of the frog's performance of its task functions. 156

My constraint: different contents for different vehicles in the same range will generally be more explanatory. 156

Soft constraint: different contents for different vehicles. 157

It follows that location is represented in the frog; numerosity (rather than *many*) in the analogue magnitude system. 157

Different components within the overall system tend to have different jobs, therefore carry different contents, although redundancy is possible. 157

(g) The appropriate amount of determinacy

We should expect more indeterminate contents in lower level systems. 157

Components have a family of relational properties in the frog case, with not enough complexity in the system to support a way of distinguishing between them. 157

This is either indeterminacy between multiple closely related contents, or a determinate content that cannot be captured precisely in natural language. 157–8

An indeterminacy for the whole system, about a collection of UE information for a collection of internal vehicles, does not imply each content is indeterminate independently of the others, since they must fit with each other in an explanation. 158

In a system with multiple interacting components, the need for a mesh between UE information carried by different components is a significant constraint on indeterminacy. 158

(h) Comparison to other theories

Like in Millikan, indeterminacy is reduced by appealing to causal explanation, which selects amongst coextensional properties, and by rejecting mediated explanations of stabilization. 158–9

My account adds: convergence with information; being more explicit about the constraint that different representations should have different contents; and the fact that some indeterminacy is a virtue in our subpersonal cases. 159

Papineau argues that indeterminacy is ruled out by the determinacy of desires, and by appealing to a component's specific function (following Neander). 159

(p.254) My account is like Price: content is constrained by explanatory role; having an information-carrying requirement; and adopting a high church view (contra Neander and Pietroski). 159–60

Price's immediacy and abstractness conditions have roughly the same effect as my call for correlations that feature in an unmediated explanation of task function performance. With multiple components, my account also favours correlations that are specific to a component. 160

Neander is the leading proponent of low church teleosemantics, tying content to conditions that an organism can discriminate between. I don't agree that contents, or long-armed functions, are restricted to things a component is responsible for on its own. 160

Neander's second argument is based on the science of how toads manage to discriminate their prey. That practice does not imply—indeed I would argue counts against—content being tied to discriminative capacities. 160–1

My account does allow that sensorily specific properties are represented in suitably articulated systems (e.g. §4.7), but that is not the case in the (stylized) frog/toad prey-capture mechanism. 161

Ryder, Martinez, and Artiga all fix contents in terms of properties that explain statistical regularities amongst worldly conditions, in different ways. 161

Objection: the property that explains co-occurrence in the world need not be the one which explains successful behaviour. 161–2

6.3 Compositionality and Non-Conceptual Representation

Some features of concepts are found in the simpler systems in our case studies: semantically significant structure, unsaturated components and (limited) generality. 162

'Non-conceptual' covers all cases outside 'concepts', which are personal level, expressed in language, and constituents of beliefs and desires. 162

Concepts obey a wide-ranging generality constraint. 162

I extend ‘saturated’ to non-conceptual representations, without semantically significant constituent structure, that have a complete correctness and/or satisfaction condition. 162

Most of our case studies do not involve predication; although many have semantically significant structure of a simpler kind, but only limited generality. 163

The plaid motion detection system (§4.7) has different layers for colour and motion. Neither has semantically significant constituent structure. 163

The PFC colour-motion case (§4.6b) and the bee dance case do have multiple components, but the components stand alone and make claims separately. They are not unsaturated. 163

These representations do have semantically significant structure, unlike the two separate representations in the plaid motion case. 163–4

Neither the PFC case nor the bee dance case involves unsaturated elements or predication. 164

(p.255) There may be unsaturated elements when place cells are used offline. 164

Our case studies do show some systematicity, and so meet a restricted, domain-specific generality constraint. 164–5

Time and place are not semantically significant features in the bee dance or PFC case (*pace* Millikan). 165

We should distinguish a different kind of ‘systematicity’, the kind found in a sign system exhibiting ‘organization’ (§5.5): there being a systematic mapping relationship from a vehicle dimension to a content dimension (e.g. more waggles = further away). 165

Concepts are reused in a wide range of contexts. Lacking that, the representations in our case studies are likely to have less determinacy. 165–6

Recap: three features of concepts are also found, to some extent, in our case studies. 166

6.4 Objection to Relying on (Historical) Functions

(a) Swampman

My accounts must overcome the standard challenges to relying on history in a theory of content. 166

This is made vivid by considering an intrinsic duplicate system with no history: a swampman. 166–7

All our case studies could have swamp duplicates. Why aren't the robust outcome functions of these systems enough to ground content? 167

Because, without appeal to history, there is nothing to undergird a distinction between success and failure of behaviour, in such systems. 167

E.g. consider a swamp system like the one in §4.7 that catches a ball, and another that just misses. 167

Neither would be making an error. Without stabilized functions, there is no room for a system that robustly produces an outcome in error. 168

The historical basis of stabilized functions allows us to make this distinction. Success is beneficial. 168

That intuitive argument is supported by the argument in Chapter 3 that selection, learning, and contribution to an organism's persistence are part of the cluster that gives explanatory purchase to representational content. 168–9

The stabilization process could be very recent, so a swamp system that starts interacting with its environment will rapidly acquire some task functions. 169

Other mental properties like memory would also build up rapidly in a swamp system. (Interactions with the environment would soon ground a substantial difference between Catcher and Misser.) 169

In sum, I cut down the scope of the challenge, since my view does not imply that swamp humans have no contents, and because content builds up rapidly during interactions; and then I make a positive argument that in the simple systems in our case studies content should be based on history in this way. 169

(p.256) (b) Comparison to Millikan and Papineau

Millikan says content-based generalizations depend on the historical kind *human*, so do not carry over to swampman. 169–70

But why don't content-based generalizations run off some currently constituted category (as other generalizations about swampman do)? 170

According to varitel semantics, that is because swamp systems do not fall in the natural cluster that underpins content-based explanation. 170

Papineau says the best a posteriori reduction of content, in our world where there are no swamp systems, is historical. 170

This undershoots: the current properties are positively less explanatorily powerful because they miss the distinctive cluster, so apply more widely and support fewer inductions. 170–1

6.5 Norms of Representation and of Function

(a) Systematic misrepresentation

For me, representing correctly, and well-functioning, are merely descriptive distinctions to which norms can be applied. That is all we should expect in these cases. 171

Even so I face the argument that, because fitness interests are sometimes best served by systematically misrepresenting what is the case, representing correctly cannot be equated with promoting fitness. 171–2

Peacocke’s example: systematically misrepresenting the predator as closer than it is, so as to run away faster, gives a selective advantage. 172

These examples typically assume that the representation in question is involved in a second pattern of behaviour, which fixes the correct content. If so, two different contents could arise on my view. 172

Fig. 6.1 The structure of the case from Peacocke (1993). 172

Our case studies do not have that kind of structure. It has not been shown that a challenge based on systematic misrepresentation is tenable in such cases. 173

A reason to think it is not is that, without further articulation, content ends up being fixed so as to align with whatever story is told about evolutionary benefit (cp. a representation theorem in decision theory). 173

Even though I reject the possibility of radical disconnection, misrepresentation and malfunction do come apart on my view, and can do so systematically (e.g. systematic false positives). 173

(b) Psychologically proprietary representation

Burge (2010) makes three arguments, on normative grounds, against teleological approaches to content. Subsection (a) covered his first argument. 174

His second argument is that they are too liberal, extending to cases where content has no real explanatory value. I argue elsewhere that my contents do have explanatory value in such cases (§2.3, §8.2). 174

Burge’s third argument is that contents should be distinctively psychological and normative. 174

(p.257) I accept that some psychological cases are more sophisticated, but some psychological cases are covered by my approach. I argue that the requirement that representation be psychologically proprietary is not well-motivated. 174–5

On normativity, Burge has a non-reductive approach, and argues that there is no need for an account in non-semantic, non-mental, and non-normative terms. 175

I agree with Burge that such an account is not required to believe that there are representations. But such an account is more illuminating, when it is available, as I argue it is here. 175

6.6 Conclusion

Varitel semantics produces less indeterminacy than informational semantics and teleosemantics. The remaining indeterminacy is what we should expect in the systems we have been considering. 175–6

The historical component of task functions is needed just to get the explanandum—successful behaviour—into the picture. But misrepresentation is not reduced to malfunction. 176

In short, varitel semantics does a reasonable job of avoiding the standard challenges in the literature. 176

Chapter 7—Descriptive and Directive Representation

7.1 Introduction

A descriptive representation is supposed to match the world; a directive representation is supposed to make a condition obtain. 177

This chapter draws the distinction in a non-theory-neutral way, within the varitel framework. 177

‘Descriptive’ and ‘directive’ are better terms than ‘indicative’ and ‘imperative’ because the latter are used in linguistics for the grammatical mood of a sentence. 177–8

Even our simple case studies may contain other modes of representing, e.g. the suppositional. 178

With beliefs and desires, the shared condition is sometimes called the content (that p), and the mode of representing is separated out, as an attitude to that content. 178

My terminology uses ‘content’ more broadly, for a full specification of representational import, including mode of representing. 178

A bodily movement can also count as a ‘condition’ C that the organism brings about. 178

7.2 An Account of the Distinction

The accounts of content in Chapters 4 and 5 are based on exploitable relations explaining task function performance, without distinguishing between relations to inputs and relations to outputs. 179

(p.258) We can supplement the accounts of content to classify the exploitable relations as playing a descriptive or directive role (or both). 179

A tempting first thought is to rely on an asymmetry in causation: caused by the world or causing an outcome. 179

But for varitel semantics the distinction should turn, not on causation, but on an asymmetry in how exploitable relations and their associated conditions figure in explaining task function performance. Descriptive representations need not be caused by their contents at all. 179

Directives are representations R where the role of the vehicle in explaining task functions depends on the fact that R produces condition C; for descriptives its explanatory role depends on C’s obtaining already at the point when the behavioural outputs prompted by R occur. 179–80

The case of corollary discharge introduces a complication which means that it is easier to define directive content first. 180

Definition: directive content (based on UE information). 180

For directive content, R’s role in explaining stabilization and/or robustness is to cause condition C to obtain. 180

Production of an outcome which is itself task functional can be explanatory as part of explaining how the whole system comes to produce this outcome in a way that was stabilized and robust (§4.2a). 180

Descriptive content concerns a condition C whose obtaining when R is tokened figures in explaining robustness and stabilization, but we need to exclude cases where the explanation is that R has the causal role to produce condition C. 180–1

Definition: descriptive content (based on UE information). 181

We don’t want the definition to imply that all directives also have descriptive content, but we do want that to be possible (e.g. in some cases of corollary discharge). 181

Where a motor command for C has a second functional role leading to behaviour whose explanation depends on condition C obtaining independently, then it should have additional descriptive content. 181

Our definition delivers that result. 181–2

For the structural correspondence case, consider a structural representation R that has a UE structural correspondence with condition $H(x_1,x_2)$. 182

Definition: directive content (based on UE structural correspondence). 182

Definition: descriptive content (based on UE structural correspondence). 182

Applied to rat navigation, offline place cell co-activation descriptively represents that location x is near to location y. 182–3

7.3 Application to Case Studies

(a) UE information

The honeybee nectar dance is a pushmi-pullyu according to my definitions of descriptive and directive content. 183

(p.259) The ALCOVE model has pushmi-pullyus at output and descriptives at the input and exemplar layers. 183

Motor programs are sometimes pushmi-pullyu, with matching directive and descriptive contents. 183

The ‘model state estimate’ in Miall and Wolpert’s (1996) predictive comparator model is in fact a pure descriptive, since it is the result of transforming the (directive) motor program into another representation. 183–4

Fig. 7.1 A predictive comparator model from Miall and Wolpert (1996). 184

The PFC colour/motion system (§4.6b) splits into some pure descriptives and some pure directives. 184–5

Pure descriptives arise in the analogue magnitude system, face recognition system and plaid motion tracking system. 185

The evidence accumulation system (§4.8) has descriptives in the cycle then directives that drive action. 185

(b) UE structural correspondence

In rat navigation, the system relies on the corresponding spatial relation obtaining. So, the correspondence has a descriptive content. 185

A directed causal graph programming a sequence of actions would have directive content. 185–6

Ryder’s SINBAD model produces cells that are tuned to sources of mutual information in the inputs it has encountered. 186

Operation of the SINBAD model can be inverted so as to use descriptive representations in directive mode to drive action. 186

So far, we just have reliance on multiple exploitable correlations, both at input and at output, not reliance on structural correspondence. 186–7

However, if the network were to make use of the fact that directed connections amongst its cells correspond to causal links in the world, then we would have a case of directive UE structural correspondence. 187

It is interesting to note that this model supports something like a content–attitude distinction. Each vehicle can be deployed in either descriptive or directive mode. 187

The rat place cell system may use place cells with directive (correlational) content to drive behaviour. 187

7.4 Comparison to Existing Accounts

This section makes comparisons with three existing accounts: standard teleosemantics, decoupling, and detecting achievement of a directive. 188

Teleosemantics: R has directive content when it has the function of producing C, and descriptive content when the producer has the function of producing R when C obtains. 188

My approach falls into the same broad camp as teleosemantics. 188

(p.260) Artiga argues that there will always be some, possibly very disjunctive, set of outputs that a representation R is supposed to produce; hence Millikan's view implies that all simple representation will have directive content. 188

My account does not have that consequence because disjunctive conditions do not generally qualify as content. 189

Price says a directive has to be a goal, something that is produced by a variety of means. Objection: that would exclude motor programs, which play an important role in how an organism calculates how to act. 189

Sterelny uses response breadth to pick out descriptives: they are not tightly coupled to any particular response. 189–90

Zollman distinguishes this way: descriptives are more tightly coupled to world states, directives are more tightly coupled to outputs. 190

My view tends to go along with decoupling in this way. But that is not the basis of the distinction. 190

A different view relies on deliberation: with directives, the consumer does not deliberate about how to act. 190

That is not a promising way to draw the distinction in our case studies because deliberation is not ever involved. 191

Third approach: a representation is directive when the organism can detect whether C obtains and stops pursuing C when it does. 191

This is too demanding for directive content in general, but where a system has this sophistication, my account implies that the representation is directive. 191

7.5 Further Sophistication

(a) More complex directive systems

This section looks briefly at four further levels of cognitive sophistication, beyond simply separating descriptive from directive representations. 192

We saw two cases where, as with beliefs and desires, the same vehicle can be used with different directions of fit. 192

In the phenomenon of secondary conditioning, a descriptive representation of condition C has come to be a directive representation causing the organism to bring about C. 192

Detecting when you have reached your goal is a further level of sophistication. The general-purpose redeployability of every desire as a belief and every belief as a desire is a further level again. 192

Many organisms have a system for sorting amongst and prioritizing potentially conflicting directive representations, including through having directives with different (and maybe variable) strengths. 192–3

So, there are at least four levels of cognitive sophistication in which descriptive and directive representations can be embedded. 193

(p.261) (b) Another mode of representing

Propositional attitudes admit of other modes of representing, e.g. supposing. There may be something like that involved when place cells are active offline. 193

Offline activation of one place cell may have a content along the lines of *suppose you were at x*. Combined with the descriptive representation that *location y is near location x*, the system concludes that *y would be nearby*. 194

So, offline place cell activity is either unsaturated or suppositional. Either way, the case introduces a kind of sophistication which might have been thought to be the preserve of propositional attitudes. 194

I remain neutral as to whether the functional role described here is the same mode of representing as the propositional attitude of supposing. 194

7.6 Conclusion

This chapter has shown how the descriptive–directive distinction can be drawn within the varitel framework. 194–5

Chapter 8—How Content Explains

8.1 Introduction

Three paragraphs highlight some distinctive features of varitel semantics and introduce the chapter sections. 197–8

8.2 How Content Explains

(a) Explanatory traction in varitel semantics

In order to see how content explains behaviour, the framework in Chapter 2 had contents as relational properties of real vehicles. 198–9

Now: do our accounts of content show how contents explain success and failure of behaviour? 199

For a primate in a numerosity experiment, picking the more numerous collection of objects constitutes success. 199

Contents are relational properties of components which thereby instantiate an algorithm for producing successful behaviour. 199

This explanatory practice applies to many cases—because evolution produces organisms in which robustly produced outcomes have been the target of stabilizing processes. 200

Representation arises when stabilization and robustness are achieved by internal workings (over vehicles bearing exploitable relations to distal objects and properties). 200

(p.262) (b) Non-semantic causal description?

Further challenge: what role is there for content, when a non-semantic causal description is always available? 200

The field of psychology looks to be full of rich content-based generalizations, but a non-semantic causal description threatens to undermine their *prima facie* explanatory force. 201

Rifle firing example: the putative semantic description marches exactly in step with the non-semantic description. 201

Varitel semantics implies that the representational explanation has vehicle–world patterns at input, and often at output, that a factorized explanation misses. 201–2

Fig. 8.1 A schematic depiction of bridging at input and output. 202

The representational explanation is based on real patterns in the way a system and its components are involved with the distal environment. 202–3

This bridging is exemplified in the analogue magnitude and rat navigation case studies. 203–4

Bodily movements considered independently of their relations to distal outcomes are often uninterpretable, e.g. the moving thumbs of someone playing a video game. 204

Bridging shows why content-based explanation breaks free from non-semantic vehicle-based explanation, allowing detailed psychological theories to have their own explanatory purchase. 204

Explanation calls for a mix of generality (breadth of application) and specificity (inductive power). Bridging delivers some generality. Specificity is offered by the rich psychological theories in which so-based contents figure. 204

(c) Doing without talk of representation

Different challenge: why not do all the explanation in terms of correlation, correspondence, and function directly? 204–5

One version: give more fine-grained explanations, just in terms of correlation, correspondence and function. That is less explanatory. 205

Second version: accept that the clusters I point to are present and the complex properties I have constructed are important. But that is to concede everything except the term ‘representation’. 205

(d) Other views about the explanatory purchase of content

Ramsey argues that representational properties can earn their explanatory keep through heuristic value or causal relevance. 205–6

Egan, Shagrir, and Burge say representational contents come in to explain how an organism performs cognitive tasks. Thus, the explanandum is already given in semantic terms. 206

Dretske (1988) has contents that are causally relevant, because they are a structuring cause of behaviour. Content based on sorting behaviour (Davies) or world-involving action (Peacocke) might also be causally relevant. 206

(p.263) Causal relevance might also be based on the compositionality of concepts, because that explains the systematicity of representational capacities or of behaviour. 206–7

Part of the explanatory purchase of content for varitel semantics is that content explains why the system is configured in the way it is and behaves the way it does (like Dretske 1988, but without the liberality). 207

I agree with Egan and Shagrir that contents allow us to see how a system can perform a task, although in my case the tasks are characterized non-semantically, not cognitively. 207

Vehicle properties also play a role, in showing how an algorithm works. This is a generalization of the point that compositionality of representational vehicles can explain systematicity of behaviour. 207

8.3 Causal Efficacy of Semantic Properties

Are semantic properties causally efficacious, or merely explanatorily relevant, on this view? 208

Varitel semantics fits Jackson and Pettit's account of why broad contents are explanatorily relevant. 208

One approach has it that both semantic properties and vehicle properties are explanatorily relevant and neither causally efficacious, with causation only in basic physics. 208

An alternative is that there can be real causal efficacy at more than one level. 208

Even if some special science properties have causal efficacy, there are further obstacles to establishing the causal efficacy of content properties. 209

My view shows why content properties are explanatorily relevant and is neutral on causal efficacy. 209

8.4 Why Require Exploitable Relations?

Are exploitable relations a necessary part of the content-constitution story? (Standard teleosemantics has no correlation requirement.) 209

Exploitable relations figure in my story because content is partly a matter of explaining how a system achieves its functions. 209

Output-only theories type representations by the type of behaviour produced; my account gives a reason to expect correctness conditions to obtain. 210

This means my account makes stronger predictions about the results of tokening a representation; but output-only teleosemantics can appeal to equivalent empirical generalizations, so the difference is not profound. 210

My account does a better job of showing why contents explain how a system performs its functions. 210

8.5 Ambit of Varitel Semantics

(a) Representation only if content is explanatory?

The complex properties my accounts depend on exist whether or not there is an observer able to make use of their explanatory potential. 210–11

(p.264) There is no requirement for contents to be explanatorily beneficial in every case where they arise. 211

E.g. they may give little or no additional explanatory purchase in a simple thermostat. 211

Robustness comes in degrees; different stabilizing processes may be co-present and align, or not. There is less explanatory purchase in the more marginal cases. 211

Where task functions are constituted by design, the other elements may be very marginal; or these can be clear cases. 211–12

In sum: representational contents are observer-independent. The explanatory utility of the representational explanations they afford will vary. 212

Finally, is my account pragmatist? It is pragmatist in the sense that content is action-derived. But it denies that starting with action guidance implies that the role of representation should be marginalized or eliminated. 212

My emphasis on the explanatory role of content suggests pragmatism of another kind. My account certainly aims to explain the discourse of representational attribution. 212

Blackburn: the pragmatist explains by mentioning not using the referring expressions of the discourse. By those lights, varitel semantics is not pragmatist. It uses the terms ‘representation’, ‘content’, etc. (and shows what they refer to). 212–13

(b) Are any cases excluded?

Does varitel semantics imply that every natural system is a representer? 213

Simple reactions to proximal inputs, e.g. plant tropisms, do not count, because there is no robust outcome function. 213

But a plant can have representational states, e.g. if it has two ways of detecting evening, closing its flowers accordingly. 213–14

Robustness in general, e.g. in cell physiology, is not robustness in the face of different inputs, so does not in general found content. 214

Internal subpersonal cases do extend beyond the psychological, to hormonal signalling and the immune system, for example. 214

Subsystems can have robust outcome functions where the ‘external’ conditions are states of other parts of the organism, but these are only task functions if there is a stabilizing process (e.g. learning processes in the cell) operating below the level of the whole organism (which is where evolution by natural selection is likely to be powerful). 214–15

The functions of hormonal signalling are probably derivative from its role in serving task functions of the whole organism, not because it has task functions as a system in its own right. Similarly for the brain. 215

Although my account is more restricted than some other theories of content, it is not restricted to the psychological. 215

The kinds of content found in personal-level cases may be restricted to the psychological, but these do not cover the subpersonal-psychological. Accounts that cover the subpersonal-psychological extend more widely. 215–16

(p.265) 8.6 Development and Content

Both with concepts and with neural networks, content often concerns the circumstances in which a vehicle develops. 216

If content is just fixed by synchronic properties, it is puzzling why there should be this connection to the circumstances of development. 216

There are empirically studied cases like that. With face recognition, new vehicles arise as a result of learning/stabilizing processes and refer to their causes. 216–217

Laurence and Margolis (2002) have a theory of natural kind concepts that links the content of a new concept to the object that caused its development. 217

That connection arises, according to my accounts, because content is fixed by features of the stabilization process (task function), and because stabilization processes often give rise to representational vehicles. 217

In sum: we can see why a new representation often represents features of the things in the environment that caused it to develop. 217

8.7 Miscellaneous Qualifications

Can contents only be about distal objects and properties? 218

A system must have distally involved task functions, so some contents must concern distal features of the environment, but it can also have representations about proximal and internal conditions. 218

A clear example of that is meta-representation, which can arise in relatively simple systems (Shea 2014c). 218

Outputs are not restricted to bodily movements and the effects of bodily movements. Chemical and electrical outputs, for example, are included. 218–19

In the case studies, vehicles have been proper parts of a mechanism. 219

In principle, properties of the whole system could interact in content-preserving ways. 219

Functional content is a supplement to informational treatments of evolutionary signalling games. It only arises because of a stabilizing process, as do my stabilized functions. 219–20

Some game-theoretic models have meaning at attractor states that are not Nash equilibria, but as attractors they can potentially form the basis of stabilized functions, hence task functions. 220

I'm neutral about where Fregean sense or mode of presentation is needed, in addition to referential content and vehicle properties. 220

Indexicality is an important issue, but I set it aside. 220

8.8 How to Find Out What Is Represented

My account is a metaphysics of content, but it has implications for how we should find out what is represented. 221

Procedure: establish behaviour that has been stabilized, consider algorithms, find which one maps onto internal processes. 221

(p.266) The role of task functions in specifying the explanandum is often implicit. The search for correlational information is explicit, but the restriction to information relevant to performing task functions is usually only implicit. 221

Information has to be carried in a way that is detectable by downstream processes. This is sometimes acknowledged. 221–2

There should be more emphasis on output correlations. Also on the circumstances that stabilize behaviour. 222

We can see why investigating illusions and errors is important. Ethology and comparative psychology are relevant to stabilization. 222

8.9 Differences at the Personal Level

How might features of the personal level make a difference to content determination? 222

Consciousness makes a difference. It could play a role in determining content; or, if it is determined by content, then various functional features of consciousness are potentially relevant to content determination (listed). 222–3

If for conscious states, or other mental states, object-level and meta-level content are fixed simultaneously, that would be relevant to content determination. 223

Relations of entailment and exclusion in a network of beliefs or concepts may play a content-constituting role. 223

For concepts, the special functional role of conscious deliberation in forming and changing our beliefs involving a concept may be relevant to content determination. 223–4

Norms applicable to belief/desire content may be interpersonal and may depend on stabilizing processes in a social group. 224

An ascriptionist theory like Dennett’s intentional stance may be the right account of content for standing beliefs. 224

Will the varitel framework, at least, still be applicable? Too early to say. 224

It is wrong to think that personal-level processes could make use of subpersonal contents directly. 224–5

Pluralism amongst different kinds of personal-level representations may not be appropriate. 225

Having answered the content question for subpersonal representation should give us optimism. 225

It gives us a fixed point from which to build. Deploying our understanding of the ways in which personal-level representations are different, understanding their nature becomes a tractable research programme. 225

However, the most important achievement, if it succeeds, is that varitel semantics allows us to understand the nature of content in subpersonal representational systems. 225–6