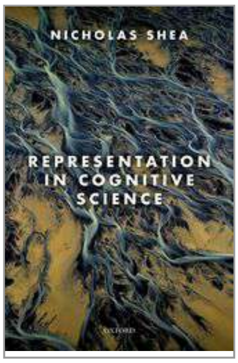


Representation in Cognitive Science
Nicholas Shea

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

Framework

Nicholas Shea

DOI:10.1093/oso/9780198812883.003.0002

Abstract and Keywords

This chapter sets out and motivates a framework for naturalizing content, ‘varitel semantics’. It is based on the explanatory advantages that flow from realism about the vehicles of content. Its distinctive features include: pluralism; eschewing representation consumers; being testing against a desideratum concerning representational explanation rather than against intuitions; and a proposal about the way exploitable relations, vehicle processing and a system’s functions come together to constitute content.

Keywords: representational realism, naturalism, vehicle of content, algorithm, externalism, homuncular functionalism, pluralism, exploitable relation, etiological function, varitel semantics

2.1 Setting Aside Some Harder Cases

This chapter sets out the framework within which I will develop my account of content (Part II). The focus is on exposition rather than a detailed defence of the approach. Nor do I mount arguments against other approaches. The framework needs to be judged by its fruits—by whether the accounts of content it underpins are satisfactory. Once that has been laid out in Part II of the book, I will return in Part III to discuss others’ views and to defend the framework developed here. I start in this section by homing in on the class of representation that forms the target of our enquiry.

Existing treatments of content have mostly started with everyday examples like occurrent beliefs and desires (doxastic states) and other conscious states. These are indeed paradigmatic cases, but they are not the only place where intentionality is mysterious and the need for a theory of content is pressing. In information-processing explanations in cognitive neuroscience the ‘information’ is in fact representational content (e.g. Franklin and Wolpert 2011, Yang and Shadlen 2007). Indeed, the cognitive sciences in general extend well beyond doxastic states and conscious states. They are shot through with representations throughout. It is perfectly legitimate for these sciences to take the existence of representations—content-bearing physical particulars—for granted. The sciences of the mind have been remarkably successful in predicting and explaining both behaviour and the goings on in the brain. That success goes a long way towards vindicating the foundational assumption that neural processes trade in mental representations. Nevertheless, the nature of representational content remains puzzling even in non-doxastic, non-conscious cases. It would be a considerable achievement to understand what makes it the case that these states have the kind of intentionality or aboutness presupposed by scientific theories.

(p.26) To do that I will set aside some features of everyday representations that make the content question more complex. Consciousness is one. I won’t deal with cases where a representation’s being conscious is relevant to fixing its content. A second feature of ordinary occurrent beliefs and desires[1] is that they enter into epistemic relations: perceptual states justifying beliefs, beliefs justifying other beliefs, and so on. I set aside cases where a representation’s entering into relations of justification for the person are relevant to its content. A third feature, which is probably related, is that they are offered as reasons in the social process of explaining to others what we believe and why we act as we do. When acting together, these reasons or verbal reports feature heavily in deliberative joint decision-making. A fourth feature is the kind of constituent structure characteristic of natural language; for example, if there were a kind of representation only available to those capable of using genuinely singular terms.

To have a suitable shorthand, I will use the term ‘subpersonal’ to cover representations for which content-determination does not depend on those complicating features: consciousness, justification for the person, a role in reason-giving interactions with other people, or being structured like natural language sentences. I am not concerned with the question of whether there is a fundamental personal–subpersonal distinction; for example, a division between incommensurate schemes of explanation (Hornsby 1997, 2000). Nor is ‘subpersonal’ supposed to connote a distinction between the whole organism and its parts. I use the term simply as an umbrella term for disclaiming these four complicating factors.

Considerable progress was made on the content question in the 1980s and 1990s. In the time since then we have learnt much more about the way mental representations are realized in neural processes. A standard line then was that we would never discern vehicles of content amongst the messy workings of the brain (Fodor 1974, 1987a). The representational level of explanation was taken to be autonomous from the neural to the extent that we should theorize about it separately, without looking to facts about neural realization as a substantial evidential constraint. More recent progress in uncovering the neural basis of representation gives the philosopher of content some fantastic material to work with: cases where a well-confirmed account of the computational processing that generates behaviour can be married with a detailed understanding of the processes in which neural representations are realized. Two cases where there is a convincing computational explanation realized in a well-supported neural mechanism are the neural mechanisms for accumulating probabilistic information about reward (Yang and Shadlen 2007) and the neural circuit responsible for motor control (Wolpert et al. 1998, Franklin and Wolpert 2011). Figure 2.1 illustrates the latter case. The details need not detain us; it is enough to notice the characteristic pattern of explanation: the circuit is described both neuro-anatomically, and also computationally—in terms of the representational contents carried by those neural areas and the way their interactions **(p.27)** compute a useful function. Because of these developments, we now have a wealth of empirical data against which to formulate and test theories of neural representation.

Since non-conscious, non-doxastic neural representations raise the problem of content in a clear way, one central aim of the book is to provide a theory of content for them. Neural representations form the subject matter of some of our central case studies. Although I'm keen to endorse the idea of neural representation and show that it makes sense, calling the book 'Representation in the Brain' would be misleadingly narrow. The same issues arise in other parts of cognitive science, where there are good reasons think representations are causally interacting physical particulars, but where their neural realization is unknown

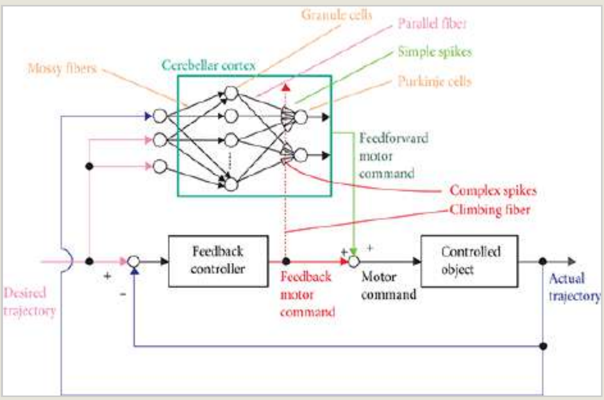


Figure 2.1 Diagram of a leading theory of motor control (Wolpert et al. 1998), used here to illustrate what is characteristic of neural representations. Prescinding from the details, neural areas are picked out

and could prove intractable. Many of these cases are also subpersonal in the sense of lacking our complicating features. So ‘Representation for Cognitive Science’ is a better description. Cognitive science does of course also encompass doxastic and conscious states, so my account is not intended to apply to all of cognitive science. What I’m aiming for is an account that applies widely within the cognitive sciences, and which underscores the legitimacy of its reliance on a notion of representation. Hence: *Representation in Cognitive Science*.

both anatomically and in terms of what is represented and computed.

My overall philosophical strategy, then, is to start with the subpersonal and work upwards. This inverts the normal approach.[2] But the normal approach has not been entirely successful to date. If we are puzzled about how there could be space in the **(p.28)** natural world for intentionality at all, then seeing how it arises in a range of cases in cognitive science will be a major step towards resolving the puzzle. Furthermore, seeing how representational content arises and earns its explanatory keep in these cases should prove a useful staging post on the way to tackling the more complex cases. So, an account of subpersonal representational content is part of a broader strategy for tackling the problem of intentionality. Given the starring role of representational notions in the cognitive sciences, it would also be a significant result in its own right.

2.2 What Should Constrain Our Theorizing?

The most widespread way to test theories of content has been to line them up against intuitions about what mental states represent. That method reached an impasse of conflicting and often theory-driven intuitions. Especially when focusing on the subpersonal, it is clear that intuitions should be given little weight. For personal-level mental states, like beliefs and desires, we have some reason to rely on our intuitions about content, our judgements about what our thoughts mean. Even there, experimental results on the unreliability of intuition when people are giving common-sense explanations of their behaviour should be make us cautious (Nisbett and Wilson 1977, Johansson et al. 2005, Carruthers 2011). When it comes to subpersonal representations, it is unclear why intuitions about their content should be reliable at all.

I take a different approach. A theory of content is answerable not to intuition, but to the role representations play in explaining behaviour. The rat finds its way back to food in a maze because it accurately represents its position and the location of food. Representing correctly explains successful behaviour and misrepresentation explains failure. A good theory of content should show how the contents it specifies are suited to explaining behaviour in that way.[3]

To do that we need to examine a range of cases where subpersonal representations explain an organism’s outputs.[4] Experimental psychology and cognitive neuroscience give us a large number to choose from. We'll mostly look at the behaviour of organisms, but artefacts like computers and control devices also produce outputs in response to their environments as a result of representation processing. I'll use 'behaviour' as a neutral term for actions and other outputs (but not internal workings) and 'system' as an umbrella term for organisms and other entities whose behaviour is representationally generated.[5]

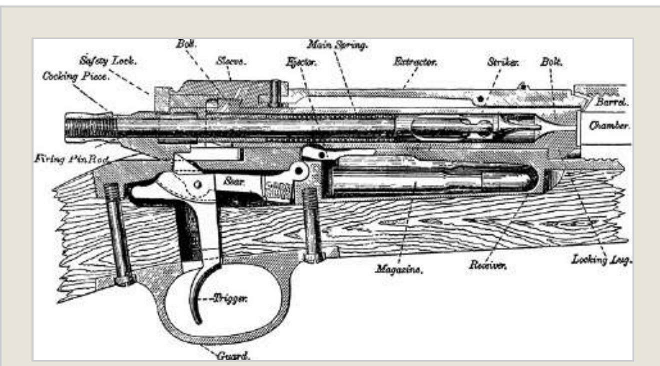
When a scientific explanation points to representational content to explain behaviour, we need to get inside that explanation to see how it works. That means getting into the **(p.29)** details of a behavioural profile and its causal underpinnings. Only with the details in view can we properly address the question: what kind of thing is representational content, to enable behaviour to be explained in that way? We should be open to finding cases where the attributed representational content is doing no work—it is just a way of understanding the system, a useful fiction—or where content is different from that attributed by the psychological theory. If our conclusions are going to have general applicability, we need to pursue a wide range of case studies so as to sample systems that have different features and operate in different ways: perceptual, motoric, and cognitive, in humans and in other animals. So, a seemingly innocuous desideratum—that representational content should be characterized by reference to its explanatory role—turns out to be a prescription for getting into the details of a wide range of case studies from subpersonal empirical psychology. That is exactly what we will do in this book.

The project is to put forward one or more theories of content that tell us how the representations involved in these case studies get their contents. Rather than intuition, our theorizing is constrained by a desideratum, something we want to explain. What we want to explain is how adverting to representational content allows us to explain behaviour. We want an account of content which shows why content plays that proprietary explanatory role:[6]

> *Desideratum*
>
> An account of how representational content is constituted in a class of systems should allow us to show why recognizing the representational properties of such systems enables better explanations of behaviour than would be available otherwise.

Since we are investigating cases where there are real vehicles of representational content, individuable non-representationally, a causal account of the operation of the system in terms of vehicle properties will always be available in principle. The vehicles in a computer are electrical currents in semi-conducting chips, which causally interact in virtue of their electrical properties; similarly for neural vehicles consisting of patterns of neural activity. Neural activity unfolds in virtue of electrical and chemical properties of neurons and synapses. Vehicle properties depend only on intrinsic physical properties of the system, its parts, and their interrelations. So, any interaction with the distal environment could in principle be ‘factorized’ into three components: the way the environment causes changes to intrinsic physical properties of inputs to the system; the way those inputs cause changes to other internal states of the system, issuing eventually in motor movements produced by the system; and the way those movements of the system cause changes to its distal environment. To meet the desideratum, **(p.30)** representational content has to offer a better explanation of behaviour than such a ‘factorized’ explanation can provide (see §8.2).

To take an example from Ramsey (2007, pp. 138, 140–1), a rifle responds to a finger movement by discharging a bullet from the muzzle. There is an internal mechanism whereby movement of the trigger (input) causes movement of the firing pin, causing ignition of the primer in the cartridge, causing explosion of the propellant, causing the bullet to travel down the barrel and exit the muzzle at speed (Figure 2.2). The movement of the firing pin is designed to correlate with movement of the trigger finger at input, and to lead to the firing of a bullet at output. A teleosemantic theory, based on deliberate design rather than evolutionary function, could treat the cartridge as a ‘consumer’ for the state of the firing pin, implying that when the pin moves it represents *the trigger has been pressed, fire a bullet.* However here a representational explanation of the behaviour of the rifle would march exactly in step with a factorized explanation that simply describes the causal chain from finger to trigger to firing pin to primer to propellant to bullet—without mentioning content at all. The widely used example of magnetotactic bacteria is also a case where a factorized explanation marches exactly in step with the putative representational explanation. As standardly described in the philosophical literature the bacterium contains a little magnet which causes the whole organism to align with the direction of the earth’s magnetic field hence to swim in that direction (Dretske 1986): magnetic field causes alignment causes direction of motion. We will see that the bacterium will not satisfy our conditions for having representational content (§8.2b).



Our desideratum is to show why representational content allows for better explanations than would be available otherwise. Theorists often make a stronger demand: that a theory of content should show why representational explanations are indispensable, **(p.31)** allowing us to explain something that could not be explained otherwise (Dennett 1971). That is too strong. If appealing to representations offers a better or more perspicuous explanation of behaviour than would be available otherwise, that is a sufficient motivation for adverting to representations when explaining behaviour. This chapter puts forward a framework designed to meet the desideratum, but it is not until we have the positive accounts of content in hand that I will be able to show how content-based explanation captures something important that the factorized explanation misses (§3.6 and §8.2b).

Figure 2.2 The mechanism for firing a rifle (an example discussed by Ramsey 2007).

I am not aiming to analyse the concept of representation: the everyday notion or the scientific notion. The theories discussed in Chapter 1 are often rejected as failing to accord with intuition. My test is whether they meet our desideratum. I take my task to be to define some technical terms and show that they are useful. Various kinds of content are defined in the chapters to come (UE information and UE structural correspondence, both based on a variety of task functions). What I aim to show is that these terms are non-empty and that the properties they pick out are useful for explaining behaviour. They are properties of internal vehicles that allow for the explanation of successful and unsuccessful behaviour of an organism in terms of correct and incorrect representation. This does not imply that an organism only has content when it would be useful to treat it as having content. That is a different kind of theory, one which makes the very existence of content relative to its explanatory merits for an interpreter. I do aim to show how content, as I define it, gets a useful explanatory role in general, but that does not imply that every instance of content is explanatorily useful, nor does it make the existence of content relative to an interpreter. The properties defined by the terms I will introduce exist whether or not anyone is around to make explanatory use of them.

2.3 Externalist Explanandum, Externalist Explanans

As we saw in the last chapter, underlying the idea that the mind processes mental representations (RTM) is a core insight: mental representations are physical particulars which interact causally in virtue of non-semantic properties (e.g. their physical form) in ways that are faithful to their semantic properties.[7] Psychological processes like thinking, perceiving, planning, reasoning and imagining consist of causal processes taking place between representations with appropriate contents. Here I explain how I think that insight is best understood.

RTM is committed to there being real, individuable vehicles of content. The problem we have just seen then becomes pressing. A complete causal account of the operation of the system will be available in non-contentful terms. Proximal stimulation at input will cause the system to undergo various internal transitions that eventuate in movements at output. Intermediate entities at various stages of this process may *have* semantic **(p.32)** properties, but content does not figure in the underlying causal story about internal transitions and bodily movements.

We can see this very clearly when we look at a computational system designed to calculate how money flows through the UK economy. Moniac is a computer that uses water in tanks to represent money. It is obvious that the various levels of water representing national income, imports, tax, and so on interact solely in virtue of the physical properties of water and of the mechanisms it flows through (Figure 2.3). The vehicles are much harder to observe in psychological cases, but the principle is the same—a non-semantic causal story is always available.

Figure 2.3 ‘Moniac’ uses water to compute the way money flows through the UK economy.

Contents come into view when we target a different explanandum. An organism interacts with the environment, bringing about distal effects in its environment, doing so by reacting to distal objects and properties in the environment. There are real patterns in the environment and an agent’s interactions with it that would be invisible if we looked only at intrinsic properties of the agent.[8] Those patterns call for explanation. They are an additional explanandum, beyond the question of how processes within an organism unfold over time—an explanandum concerning the organism’s interactions with its environment. Given an externalist explanandum, externalist properties of the system and its components are good candidates to figure in the explanation (Peacocke 1993). Which externalist properties? Plausibly content properties, if contents are externalist. Representational contents figure in explanations of how an organism interacts with its environment and achieves distal effects in its environment. So it makes sense that content should be externalist, determined partly by extrinsic properties of representational vehicles (cp. Ramsey 2007, pp. 95–6). Contents would then be suited to explaining real patterns that are the result of organism–environment interactions. They would then explain things which purely intrinsic properties of the system do not.

Not every possible interaction between a system and the environments it could be placed in calls for explanation in this way. The way a spider is swept down a river does not. Nor does every system enter into the sorts of interaction that representations are characteristically called upon to explain (none of a river's interactions call for the river to be a representer). I take from teleosemantics the idea that it is when a system is performing a function (e.g. getting nectar for the hive) that representational explanation becomes pertinent. Or at least, that explaining the performance of functions is one central way that representation gets its explanatory purchase. Chapter 3 develops a more general account, going beyond evolution by natural selection, of which input–output behaviour generated by an organism counts as functional. I call these 'task functions'. The important point for us now is that an organism with task functions achieves distal effects in its environment, can do so successfully or unsuccessfully, **(p.33)** **(p.34)** and does so by reacting to and interacting with distal objects and properties in its environment. A task function performed by an organism comprises an explanandum, about an organism's interactions with its environment, to which representational explanations can be addressed.

A task function is mapping from some states to others. For example, a rat can take itself from a range of locations to a new location where there is food and ingest it. A person can take some materials and make them into a coat. The first is a mapping from one organism-environment relation to another, the second a mapping from one state of the environment to another. Both are mediated by the activity of the organism. For a given function, there are many possible ways to perform it. Doing arithmetic, people perform a function that takes written numbers as input (say) and produces a written number, their product, as output. There are many different ways of achieving that input–output mapping. I might use long multiplication in columns with carrying. That is a way of producing the appropriate input–output mapping. The same is true for moving from one location to another or transforming some materials into a coat. Those mappings between states can be achieved in multiple ways, and there is a fact of the matter about how a particular organism does it.

David Marr's algorithmic level specifies the way a function is carried out by a system (Marr 1982). Long multiplication in columns is an algorithm for multiplying together any two numbers. Stretching the term slightly, I am going to use the term 'algorithm' for the way an organism carries out functions of the kind I have just described (navigating its environment, getting food, making tools, etc.). In the multiplication case the organism is calculating the input–output function (from two numbers to their product), but in the other cases the idea is looser. The organism carries out an algorithmic process over representations in a way that leads to it achieving the function (e.g. getting from one location to another and ingesting the food there). In my sense an algorithm is a sequence of operations between representations that leads to an organism performing a function.[9] The sequence of operations is a computation carried out by the organism.[10]

In my usage algorithms are concrete. They are a way of processing representations that is realized in some organism or other system. An algorithm described in terms of transitions between contents is neutral about how those transitions are realized, save that there must be states of the system that carry the appropriate contents and undergo the **(p.35)** appropriate transitions. Those transitions must also be describable non-semantically, in terms of the way the system moves between physical states—what is often called a syntactic description. It is this constraint which makes for a realist answer to the question: which algorithm is system S using to perform function F? Steps of the algorithm have to map onto causal transitions in the internal processing going on within the system.[11] There are several ways the problem of achieving a given input–output mapping could be broken down into intermediate steps (see Figure 2.4). A given system uses one of them. Vehicle realism means that the representational explanation of behaviour is an account of the particular way a system achieves a given input–output mapping, hence of how it manages to perform a task function.

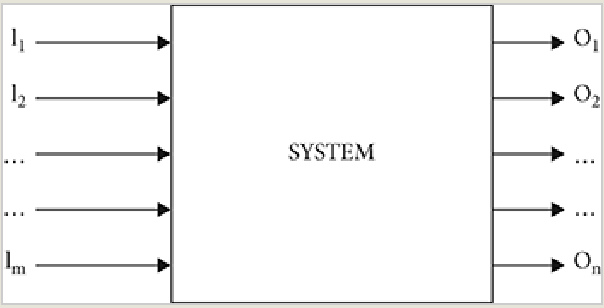


Figure 2.4 The input–output mapping produced by a system does not fix what goes on inside the box. The mapping could be generated by a look-up table and usually also by several other algorithms.

In most examples from the cognitive sciences the inputs that the system responds to and the outputs being explained are not intrinsic features of the organism, but are partly externalist.[12] Suppose we have a system that has been trained to track the direction of motion of surfaces and make a movement in the corresponding direction. One algorithm for doing that keeps track separately of the colour and motion of small portions of surface, and then combines that information to infer which portions of surface are parts of the same surface and what the overall direction of motion of each surface is (§4.7). Steps of the algorithm are described in terms of their representational content, such as representing the colour of the surface-portion at such-and-such location. Processing a series of representations with those contents is the way the system produces the distal input-output mapping.

As we just saw, if contents are going to explain how a system performs a distal function, we should expect content to be determined in part by extrinsic properties of the vehicles of content: relations those vehicles bear to objects and properties outside the system. Which relations? Here I draw on Peter Godfrey-Smith's idea that representations bear exploitable relations to features of the environment (Godfrey-Smith 2006). Godfrey-Smith takes that to be part of his 'basic representationalist model', but the **(p.36)** idea is still applicable when we drop the requirement for a representation consumer (§1.5). The entire system processes a variety of representations, possibly in complex ways, so as to carry out a distal task function from some states to others. In order to perform that function, the system makes use of the fact that intermediate components—vehicles of content—bear exploitable relations to distal features of the environment. One useful relation is to have a component that correlates with a relevant feature of the environment (Chapter 4); for example, that correlates with the colour of part of an object. Another useful relation is to have a structured set of components whose structure mirrors an important structure in the environment (Chapter 5); for example, to have a cognitive map of the spatial environment. The system as a whole makes use of these exploitable relations in calculating how to behave.

To be an implementation of an algorithm for performing a system's distal functions, internal components must have two kinds of properties at once. Causal transitions between vehicles must be those called for by the algorithm. That is a matter of the intrinsic properties that drive internal processing. And components must also have extrinsic properties that give rise to the contents which are called for by the algorithm. (How that can be so is the subject of the rest of the book.) Those contents must be respected when the vehicles are processed: intrinsic properties of the vehicles and the system in which they are processed must be such that the transitions between vehicles make sense in the light of the vehicles' relevant extrinsic properties. Exploitable relations are the link between internal components and the distally characterized task function which the organism is performing. It is the coming together of extrinsic properties and intrinsic properties in this way that gives rise to content.[13]

2.4 Representation Without a Homunculus

One alluring but mistaken way of thinking about mental representations is as being inner sentences that are understood by some internal homunculus. The model is the way we understand external sentences: we hear the words and attach meaning to them. It is a mistake to think we do something similar with mental representations: that when they occur in the mind we have to look up their meaning before we can reason or act appropriately. That would require some kind of inner interpreter of the mental representation, launching a regress.

The 'homuncular functionalist strategy' avoids the regress (Dennett 1978). First off let's see how this works if we presuppose representation consumers, before generalizing the insight to my account. The consumer of a representation does not understand its meaning. It is just disposed to react to a representation by producing a certain behaviour. The consumer is not acting that way because the representation has a certain meaning. Rather, the fact that the consumer behaves in that way constitutes the representation as **(p.37)** having a certain meaning. Consumer honeybees don't need to understand the dances they observe; they just need a causal disposition to fly off to the corresponding location. Their behaviour constitutes a dance as having a certain meaning.

The homuncular functionalist strategy is to show that a complex mental capacity arises from the interaction of simpler components, bottoming out in components whose causal operation does not presuppose any mentality or intentionality. My account of content (Chapters 3–5) does not rely on content-constituting consumers. But it still makes use of the homuncular functionalist strategy. Content arises out of the fact that a system has a certain kind of internal organization and is performing a certain function. Nothing in the system needs to interpret the internal representations or understand their content. Later chapters contain detailed proposals for the way content arises when functions, internal processing, and exploitable relations come together in the right way. These are all perfectly un-mysterious natural facts about an organism, computer or other system. The system's interactions with its environment have stabilized the performance of certain functions. It has a certain internal organization. Components correlate with or correspond structurally with distal features of the environment. If content is constructed out of these properties, as I claim, then content properties arise automatically when a system of the right kind is embedded in the right kind of task environment. They don't presuppose an internal understander of representational content.[14]

2.5 What Vehicle Realism Buys

In cases where realism about representation is justified there is a further question about whether it offers any explanatory advantages. This section offers a perspective on that question which fits with the account to come.[15]

According to RTM, transitions between mental representations are faithful to their contents, amounting to an algorithm by which a system achieves its functions. It is not just by chance, or through a look-up table, that the system instantiates a useful input–output mapping. It implements an algorithm and, if more than one algorithm would produce the same mapping, there is a fact of the matter about which one the system is in fact using. That is underpinned by the commitment to there being real vehicles of content. The algorithm must operate over a set of vehicles that can be individuated non-semantically and must follow a series of processing steps that can be specified non-semantically. Thus, vehicle realism is needed for representations to be involved in explaining how a system achieves its task functions in the way discussed above (§2.3). **(p.38)** That is the first of three explanatory advantages of vehicle realism: phenomena that it allows us to explain in a distinctive way.

The second is that it predicts a pattern in the way errors propagate. Not only does correct representation explain success and misrepresentation explain failure, but we can make predictions about patterns of failure. An incorrect representation will produce consequences in downstream processing: one error will propagate and lead to other errors downstream, errors that make sense in the light of the misrepresented content. Correlatively, parts of processing that occur before the misrepresentation, or are insulated from it, will not be driven into error thereby. Consider a mechanism that computes object motion by first representing the colour and local motion of small portions of a surface, and then integrating that information into representations of moving surfaces. A misrepresentation at an early stage, such as about local colour, is likely to lead to an error about global motion at a later processing stage. The converse is not true (in a case where only feedforward processing is involved): an error introduced at the stage of computing global motion will not cause errors at earlier stages, such as representing local colour. Ascriptionist views about content do not predict these kinds of systematic relations between incorrect representations.[16] If representations were not real entities in the system, individuable non-semantically, we would lack a ready explanation of why errors should propagate and pattern in these ways.

Thirdly, vehicle realism explains a familiar pattern of stability and change in representational capacities over time. A system tends to keep the same representational resources over time and, when they change, representations tend to be gained and lost piecemeal. Exploring an environment, we learn about new locations one at a time. If representational contents were just attributed as a useful summary of behavioural patterns, it would be unclear why changes to the system's behavioural dispositions should go along with piecemeal rather than wholescale changes to the ascribable contents. In cases where this phenomenon is observed empirically, the representational realist has a ready explanation in terms of the gain and loss of representational vehicles.

Those three patterns of explanation depend on realism about representation: upon there being a substantial non-semantic sense in which an individual token counts as being the same representation again.[17] What makes different tokens instances of the same representation is that their non-semantic properties ensure they are processed in the same way by the system. So representational vehicles can be individuated non-semantically, in terms of intrinsic properties of the system.

(p.39) Care is needed with the idea of a vehicle. A vehicle is the particular that bears representational content. Words on the page are vehicles of content. In referring to the word we don't just pick out ink marks on the page, we individuate them as falling under a type. The word 'barn' is a type that includes 'BARN', '*barn*' and 'barn'. However, the way those marks get their meaning depends not just on their intrinsic shape, but on what language they're being used in. The string of letters 'barn' in Swedish means *child* not *barn*. I will use the term 'syntactic type'[18] for the way of typing vehicles that aligns with content assignment: same syntactic type ensures same content.[19] For a case like 'barn' we would commonly say that the same representational vehicle means different things in English and Swedish. So, vehicles are not the same thing as syntactic types. The same vehicle can fall under different syntactic types in different contexts. Syntactic typing depends on the way a vehicle is processed. Analogously, the vehicle 'barn' is processed one way in English and another in Swedish. We are not looking at natural language and in our cases the way a vehicle is processed depends only on intrinsic properties of the organism/system doing the processing. So, although syntactic type need not be an intrinsic property of the representational vehicle, syntactic types can be individuated in terms of intrinsic properties of the system.

In short, vehicles are individual bearers of content picked out in terms of intrinsic processing-relevant non-semantic properties; and syntactic types are ways of typing vehicles into non-semantic types that are processed the same way by the system, and so are guaranteed to have the same content. In the brain, a distributed pattern of firing in a cortical layer can be a vehicle of content. Neural reuse means that the same pattern of firing may be put to different uses and processed differently when the organism is performing different tasks. So, the same neural vehicle (pattern of firing) may fall under different syntactic types as its effective functional connectivity changes. It may represent spatial location when processing is set up one way, and episodes in the organism's past when processing is set up another way.

Recall the dual nature of content (§2.3). Content arises from the convergence between an externally specified function being performed by a system and internal processing which implements an algorithm for performing that function. It follows that whether an internally specified state counts as a vehicle of content at all depends in part on the system's environment. Being a representation is not dependent only on intrinsic properties of the system. Syntactic typing is therefore partly externalist. In Shea (2013b, pp. 504–7) I give an example where what counts as one vehicle for one task divides up into many vehicles when the system is performing a different task. Is syntactic externalism compatible with the explanatory advantages which I have just claimed follow from vehicle realism? Yes, because it still follows that instances of the **(p.40)** same syntactic type *within a given system* will share processing-relevant intrinsic properties. That is what is needed to secure the advantages of realism: to give reality to the algorithm, to predict relations between errors within a system, and to explain stability and piecemeal change of representational resources in a system over time. Which intrinsic properties count as syntactic types in a given system will however depend on factors extrinsic to the system.[20]

Syntactic types can be based on properties of dynamical processes. Indeed, neural firing rate is a dynamical property. Dynamical systems theory is used to launch many different arguments against the representational theory of mind but, taken alone, the observation that dynamical processes are responsible for generating behaviour does not in itself undermine representationalism. Elements in a dynamical system can have vehicle properties that are calculated over so as to implement an algorithm for producing appropriate behaviour. To take an imaginary example, suppose walking depends on synchronizing the frequency of two dynamic loops, one involving each leg, the brain involved in both. The oscillation frequency of one leg-involving loop is not fixed by properties of motor neurons alone. It also depends on the weight of the leg, the physical properties of the bones and muscles, how they are coupled to each other, and their coupling to the brain via outgoing and incoming nerves. The phase offset between the oscillations in the loops for the right and left legs could be a vehicle of content; for example, an imperative representation of whether to speed up or slow down. It could interact in internal processing with other dynamic vehicles; for example, the recent rate of energy depletion (acting as a representation of urgency). There, dynamical properties would interact in ways that are faithful to representational contents.

It is of course a substantial question whether a dynamical system is a representational system and whether any dynamical properties are the basis for syntactic types. Useful behaviour can be achieved by some appropriately organized dynamical systems without any representations being involved at all. However, our framework applies very readily to dynamical cases and there is nothing in dynamicism as such which counts against dynamical properties being vehicles of content. Dynamical parameters like frequency, resonance, phase, impedance, and gain are all candidates.

I end this discussion of vehicle realism with a brief note about the underlying metaphysics and its relation to reductive and non-reductive physicalism. One way of naturalizing content is by reducing it to something else: finding a property identity. On the reductive view, having the representational content p is identical to having some (possibly complex) non-semantic, non-mental, non-normative property. That would indeed show, in naturalistic terms, how content is determined. A naturalistic theory of content need not, however, be reductive. It is a familiar point that many special sciences **(p.41)** deal in properties that are not reducible to a more fundamental science. That is likely to be true of representational content as well.

Non-reductive physicalism is compatible with there being exceptions to the generalizations which connect properties in different domains, with there being *ceteris paribus* bridge laws between different schemes of explanation. So, the account which says how other properties determine content properties could admit of exceptions, provided content supervenes globally on physical properties.[21] A sufficient condition for content determination, although it has nomological force (it is a non-accidental generalization), may admit of exceptions where the condition is satisfied but there is no content, exceptions that can only be explained at some other more fundamental level.

Furthermore, it would be adequate to have a series of different content-determining conditions. Each would be a content-determining sufficient condition[22] applicable to certain cases. That would be enough to show how an appropriate constellation of properties from other special sciences gives rise to content. There is no need to find a single set of necessary and sufficient conditions that covers all possible cases. Naturalism is a substantial requirement but it does not demand that we find a property identity.[23]

2.6 Pluralism: Varitel Semantics

So far I have set out a framework for realism about mental representation. The framework has two variable elements: the source of the distal functions performed by a system; and the nature of the relations that elements in the system bear to the environment, which are exploited in order for the system to perform those functions. The case studies to follow will show how these arise in different ways.

Two types of exploitable relation cover all the cases we will consider: carrying correlational information (Chapter 4) and bearing a structural correspondence (Chapter 5). I am not committed to there being a single overarching property, bearing an exploitable relation, which covers both of these cases without over-generalizing, applying to too many other cases. I don't take it to be an objection to the account that it offers more than one sufficient content-determining condition. So, I present the accounts disjunctively. Exploited correlational information appears in one set of sufficient conditions (conditions for 'UE information', Chapter 4), exploited structural correspondence appears in another (conditions for 'UE structural correspondence', **(p.42)** Chapter 5). I am not attempting to define a single overarching technical term that covers both cases. If a definition that would cover both cases applies beyond cases of correlation or structural correspondence, then there is a significant risk that the resulting notion would apply to too many cases. Liberality per se is not objectionable, if it is just generality, but liberality is an objection if it robs content of its distinctive explanatory purchase. Therefore, I carry out the project in a way that is open to pluralism: to content being constituted differently in different cases.[24]

Functions are a second source of pluralism within my framework. Different kinds of function can underpin content. I have already suggested the idea that stabilizing processes other than natural selection can underpin a distinction between successful and unsuccessful behaviour (§1.5). Dretske's case of instrumental conditioning is one example (Dretske 1988). The next chapter argues that at least four different processes give rise to teleofunctions: evolution by natural selection, behavioural learning from feedback, contribution to persistence of an individual organism, and deliberate design. We can recognize that several different processes give rise to teleofunctions without being committed to there being a single overarching theory of function that covers all of the cases—without over-generating and hence robbing the category of its explanatory purchase.

I unite three of these teleofunctional processes under the label 'stabilized function' (§3.4), and all four under the label 'task function' (§3.5). This makes it seem as if I do have a single overarching account of function: task function. In fact, the label 'task function' is just a terminological convenience. Since stabilized functions and task functions have disjunctive definitions, they in effect generate a series of different conditions for content. That is a second source of pluralism, giving us 2 (exploitable relations) x 4 (functions) content-determining conditions. Those conditions bear a striking family resemblance to each other, but I am not committed to there being a single ur-theory which encompasses them all without being too liberal (i.e. which still ensures that there is something distinctive about content explanation). These are all properties that deserve the label 'representational content', but the result of pluralism is that I am not offering a single overarching set of necessary and sufficient conditions for content.

A final source of pluralism is the simplifying move at the start of this chapter: setting aside representations at the personal level. I do think we will need a different theory to account for the content of beliefs and desires, and of conscious states; maybe more than one. But I don't need to make an argument to that effect here. For now my claim is that content may be constituted differently at the personal level, so the accounts offered below should not be tested against the contents of personal-level states.

(p.43) I will not be making a positive argument for pluralism. The point of being open to pluralism is that it allows me to resist the pressure to find a single overarching necessary and sufficient condition that covers all the cases. We may get one theory of content that gives us a good account of the correctness conditions involved in animal signalling, say, and another one for cognitive maps in the rat hippocampus. There is no need to find a single account that covers both.

When in the past I argued that a theory of content which combines correlational information with teleofunctions is applicable to some simple cases like animal signalling I gave the combination the label 'infotel' semantics (Shea 2007b). The framework developed here, as well as being different in some respects, is also more widely applicable. A variety of exploitable relations are involved: correlational information and structural correspondence. Indeed, it could turn out that other kinds of exploitable relation exist in other cases. A second source of variation is the range of different functions that can underpin content. So 'varitel' semantics seems like a good term, with the 'vari' marking both variations in the exploitable relations and variations in the teleofunctions involved. The resonance with 'varietal' is apposite, registering the fact that my account of content comes in several non-overlapping varieties.[25]

This chapter has set out the framework of varitel semantics and motivated it as an approach to naturalizing content. It has several distinctive features. Pluralism is one, as is the focus on what I have been calling subpersonal contents. My account does not rely on a representation consumer that plays a content-constituting role. Eschewing intuitions about cases is also a move away from the earlier literature. Although looking at the explanatory role of representation is not new, the desideratum set out above is somewhat distinctive. I also offer my own particular take on realism about representations and its explanatory advantages; and on exploitable relations and the dual nature of content. That is the prospectus. We move now to details of the positive accounts (Chapters 3–5). **(p.44)**

Notes:

([1]) Standing beliefs and desires are not conscious, but it is not clear whether there are representational vehicles for these contents.

([2]) Cf. Karen Neander's recent book which also focuses on simpler cases. Her target is non-conceptual representations (Neander 2017, pp. 27–46).

([3]) I remain neutral on whether representation has further explanatory roles; for example, explaining why internal processing unfolds in a certain way.

([4]) In all of our case studies, outputs are actions and their consequences; but other kinds of output can also be considered; for instance, physiological, hormonal, and neurochemical outputs.

([5]) This includes some systems that have organisms as subsystems. The honeybee colony is a system in this sense. An individual consumer bee is a subsystem of this system.

([6]) I call this a desideratum, rather than a necessary condition on the existence of content. If it were not met, it's not clear that we would be forced to give up on there being representational content, rather than changing our expectations about the nature of content.

([7]) §8.3 discusses the causal efficacy of semantic properties.

([8]) I adopt Dennett's catchy terminology without aiming to capture exactly what he meant by 'real patterns' (Dennett 1991). For me, real patterns are observer-independent approximate regularities occurring at some level of description whose existence allows us to describe the system in a more compact way at a more coarse-grained, less fundamental level (cp. Ladyman and Ross 2007, Ladyman 2017). Being more fundamental is a matter of applying at more scales (of length, time, or energy).

([9]) The stipulation that algorithms must involve representations is just for convenience of exposition. I am not assuming a semantic view of computation individuation. A sequence of operations over non-semantic states that can be carried out in a finite amount of space and time could also be described as algorithmic. On some views it could count as a computation irrespective of whether anything is represented. In that sense the rules for the flow of activity in a connectionist network would count as algorithmic, as would the learning rule. Which kinds of connectionist processing count as algorithmic in my sense depends on how it is appropriate to describe them representationally (Shea 2007a).

([10]) Some theorists reserve the term ‘computation’ for processes that deal in discrete states (Eliasmith 2010, p. 314), whereas others use it more broadly, so that the idea of analogue computation is not self-contradictory. I adopt the broader usage, covering all cases where representations are physical particulars, processed in virtue of their vehicle properties in ways that respect their semantics.

([11]) The processing has to undergo transitions that are called for by the algorithm, hence appropriate to the contents represented, but that does not imply that the causal processing is sensitive to content.

([12]) That feature is not widely noted in cognitive science. It is debated in philosophy (Egan 1991, Segal 1991).

([13]) … gives rise to the kind of content we are investigating here (see point about pluralism below). That caveat is implicit throughout.

([14]) Nor does content depend on an interpreter in a second sense: an external interpreter that treats the system as having contents. Having content of the kind described here depends on having a certain complex of observer-independent properties. Systems that have these properties are susceptible to a special scheme of explanation, but being so-explicable is not what makes it the case that a system has contentful representations: see §4.2b and §8.5a.

([15]) I return to these issues in Chapter 8.

([16]) NB Davidson’s and Dennett’s views were not intended to apply to subpersonal representations.

([17]) Neo-Fregean senses would explain some phenomena of the second and third type (patterns of error and piecemeal change). (Recall that I'm leaving aside contents at the level of sense, if they exist, and just focusing on referential content.) Senses won't replace vehicles. Vehicle realism is still needed to secure the first explanatory advantage. It is also needed to explain differences that go beyond modes of presentation, e.g. between different people who grasp the same sense, or a single thinker failing to identify two representation tokens which have the same sense. I don't here get into the converse question of whether the idea of vehicles and syntactic types allows us to do without senses (Millikan 2000, Sainsbury and Tye 2007, Recanati 2012).

([18]) Syntactic does not here imply that the representation has to have constituent structure. It connotes that aspect of syntax that is about individuating the content-bearers, doing so non-semantically. Given the problems with the term 'vehicle' there seems no better term than 'syntactic type' for the non-semantically individuated typing to which contents are assigned.

([19]) I am leaving aside problems of indexicality (§8.6).

([20]) Oron Shagrir makes the same argument about the nature of computation (Shagrir 2001), where similar issues arise. On his view (a version of the sematic view of computation), whether a system is performing a computation depends in part on factors extrinsic to the system (Crane 1990, Bontley 1998, Horowitz 2007).

([21]) That is, there should be no content difference without a non-semantic, non-mental, non-normative difference somewhere. A case which counts as an exception to a *ceteris paribus* bridge law should be different in some respect to those which fall under the law.

([22]) While aiming only at a sufficient condition, we still want to avoid otiose clauses or unnecessary requirements. Every requirement should be a necessary part of the sufficient condition.

([23]) A disjunction of such conditions, even if the list were closed, does not automatically amount to a reduction, since arbitrary disjunctions of properties may not be the right kind of thing to figure in a reductive property identity.

([24]) My pluralism was inspired by Godfrey-Smith (2004), although his pluralism about representation in cognitive science is based on variation in what scientists think the most fundamental basis of meaning is when they apply the basic representationalist model. My pluralism has a different motivation.

([25]) My neologism sounds in my head closer to 'vary-tel', a near-rhyme of 'fairy-tale' (although I'm hoping my account is not one of those).