Teleosemantic Modeling of Cognitive Representations

Marc Artiga
(forthcoming in *Biology and Philosophy*)

Abstract

Naturalistic theories of representation seek to specify the conditions that must be met for an entity to represent another entity. Although these approaches have been relatively successful in certain areas, such as communication theory or genetics, many doubt that they can be employed to naturalize complex cognitive representations. In this essay I identify some of the difficulties for developing a teleosemantic theory of cognitive representations and provide a strategy for accommodating them: to look into models of signaling in evolutionary game theory. I show how these models can be used to formulate teleosemantics and expand it in new directions.

1. Introduction

A perplexity underlies much work in cognitive science. On the one hand, classical models of cognition explain human capacities by attributing representations. Indeed, it has often been suggested that the appeal to representations is one of the marks of contemporary cognitive science (Sternberg, 2009) and might be unavoidable (Sprevak, 2013). On the other hand, however, we still lack a general theory spelling out what representations are. There is just no widely accepted naturalistic theory of content for cognitive representations. Although some proposals have been made, none of them has reached widespread agreement. This is puzzling. The main goal of this paper is to suggest a plausible avenue for addressing this worry. In a nutshell, I will argue that a particular way of developing teleosemantics can provide a naturalistic theory of cognitive representations.

Teleosemantics is probably the most popular naturalistic theory of content. One reason is that it has yielded promising results in some fields, such as communication theory (Allen and Bekoff, 1999; Stegmann, 2009, 2013) or genetics (Godfrey-Smith, 2000; Maynard Smith, 2000; Shea, 2007). Nonetheless, the long-standing debate on this approach has also uncovered some important limitations. These shortcomings are specially pressing in the context of cognitive representations, since these states possess certain features that are not shared by other types of representations. First of all, cognitive systems have an overwhelming connective complexity. Secondly, many representations are not directly connected to any environmental input or behavioral output. Thirdly, it is now generally accepted that in many cases information travels in multiple directions, not only bottom-up, but also top-down and horizontally. As we will see it is not obvious how all these features fit into the traditional sender-receiver model assumed by teleosemantics. Furthermore, I will argue that these specific difficulties derive from a more general worry: presently we do not know how to extend the basic teleosemantic framework to accommodate more complex representational mechanisms. Accordingly, the defense of a teleosemantic theory of cognitive representations is not just an application of a pre-established theory. We need to show how these intricate structures can be accounted for, which requires a new interpretation and development of

this classic naturalistic approach.

In this essay I will argue that recent models in evolutionary game theory can be used in order to overcome the limitations of the standard teleosemantic theory. Whereas most people working within this area tend to reject teleosemantics and endorse some sort of informational semantics, I will show that these models can also be used to describe and extend the teleosemantic framework. Moreover, I will defend the claim that once we adopt this perspective, cognitive representations can plausibly be naturalized. The sort of questions posed by cognitive representations can be answered by looking into evolutionary models.

The paper is structured in two main parts. In the first one I will put forward the teleosemantic theory I will be relying on and I will argue it has significant shortcomings, as highlighted by cognitive representations. More precisely, I will present three specific difficulties (the existence of multiple connectivity, isolation and feedback loops) and a general problem: at the moment we do not know how to extend the basic teleosemantic framework to complex representational mechanisms. In the second part of the paper I will put forward a suggestion for dealing with these difficulties. First, I will argue that certain models in evolutionary game theory can be used in order to describe teleosemantics more formally. Secondly, I will show that these tools can be employed to extend teleosemantics in new directions. Crucially, some of these new teleosemantic models can accommodate the cognitive structures that were problematic in previous accounts. As a result, I will suggest that teleosemantics might offer a plausible naturalistic theory of content, which might also unify cognitive representations and other kinds of representations within the same paradigm.

2. The Teleosemantic Project

Teleosemantics is a naturalistic theory of representation. Naturalistic theories of representation seek to specify the conditions that must be met for an entity to represent another entity. To properly qualify as 'naturalistic' in the intended sense, the entities mentioned in the *explanans* must satisfy two conditions: they have to be non-representational and must be the kind of entities postulated by some science. For example, well-known naturalistic theories of representation appeal to causality (Stampe, 1977), information (Dretske, 1981; Usher, 2001) or laws (Fodor, 1990). In this essay, I will concentrate on a promising approach in this naturalistic tradition: Sender-Receiver Teleosemantics (Millikan 1984; Godfrey-Smith, 1996; Papineau, 1987; Shea, 2007).

The distinctive feature of Sender-Receiver Teleosemantics (henceforth, 'SR-Teleosemantics') is that it analyzes representations by appealing to two notions: sender-receiver systems and functions. First of all, a *sender-receiver system* is composed of two mechanisms; the first one delivers a set of outputs given certain inputs, and the second one takes the first mechanism's outputs as inputs and delivers a further output. Thus, a sender-receiver structure simply consists of two coupled mechanisms in which the output of the first mechanism serves as input for the second one. The other concept playing a pivotal role in the teleosemantic analysis of representation is the biological notion of *function*. In particular, according to the etiological definition of function widely used in biology, the function of a trait is the effect of past instances of the same type that explain why it was selected for. For instance, the function of the pancreas is to produce insulin and other hormones because this is the effect that explains why it was selected for (and, ultimately, why current organisms have this organ). Functions are selected effects (Millikan, 1989; Neander, 1991).

The main strategy of SR-Teleosemantics is to combine the notions of sender-receiver structure and function in order to provide a naturalistic theory of representations. More precisely, representations are states produced in sender-receiver structures, in which both sender and receiver have certain functions. The function of the sender is to produce a state (the representation) when certain worldly

affairs hold, and the function of the consumer is to act in certain ways when these states are produced. To be a representation is just to be an intermediate state in such a structure. More precisely, given the sets of states $M = \{M_1, M_2, ...M_n\}$, $S = \{S_1, S_2, ...S_n\}$ and $A = \{A_1, A_2, ...A_n\}$:

SR-TEL:

Mm^1 is a representation iff[2]:

1. Mm is an intermediate state within a sender-receiver system.
2. The sender's function is to emit Mm when a state S_m is the case.
3. The receiver's function is to produce A_m when Mm is emitted.

For instance, meerkats (*Suricata suricatta*) produce different alarm calls (M_1, M_2, M_3) when there is a mammalian predator (S_1), a snake (S_2) or a bird (S_2) (Manser et al. 2002). Since we might plausibly suppose that the mechanisms responsible for emitting these signals and responding to them have been selected for, they fit the definition provided above. Accordingly, they constitute representations.

SR-Teleosemantics not only provides an account of what representations are, but it also explains what determines their content. On a popular approach deriving from Millikan (1984), the feature represented (the state of the world a signal is supposed to map onto) is the worldly affair that has historically explained why the receiver has fulfilled its functions. In other words, the content of each signal is determined by the feature that explains why the corresponding behavior contributed to survival. In meerkats, each of the signals (M_1, M_2, M_3) elicits a different response, which is adequate for a different type of predator. Behaviors elicited by M_2, for instance, provided an advantage to meerkats because they enabled them to escape from snakes (S_2). For this reason, M_2 represents something like *there is a snake around* (i.e. S_2 holds).

A key aspect of this theory that accounts for its great explanatory power is that in principle there is no restriction concerning the kinds of mechanism that can instantiate this framework. In the previous example different individuals play the role of sender and receiver, but a single organism can also possess various internal mechanisms that can be considered senders and receivers.[3] A classical example is provided by the prey-catching mechanism of anurans. Whenever they detect a black thing moving in a certain direction at a specific velocity (which, often enough, is an edible bug) they throw the tongue out to catch it. This configuration clearly illustrates the previous schema: there is a sender (perceptual mechanism), a receiver (the motor system) and some internal state that qualifies as a representation.

Meerkats and frogs exemplify the kind of cases teleosemanticists have focused on so far. There is of course an open debate as to whether the explanations provided in these cases are satisfying, but in this paper I would like to leave these questions aside. In a sense, our question is more fundamental: can this approach be extended to more complex mechanisms, such as human cognitive states? Until recently, few have attempted to address this question directly. This is unfortunate. Anyone interested in the naturalization of representations should be troubled by the fact that, after more than thirty

---

1 'Mm' refers to a type of state. Tokens will count as representations in virtue of belonging to that type.

2 Even though I state the account in terms of necessary and sufficient conditions, SR-TEL is a first approximation to teleosemantic theories, so probably some of these conditions might need to be slighlty modified and some others might need to be added (the second part of the essay will partially examine this issue).

3 Indeed, some cases might lie between these two interpretations. Honeybees, for instance, use a well-known set of communication signals to inform fellow bees about a wide range of features and it is debatable whether a honeybee colony should be considered a collection of organisms or a single superorganism (Godfrey-Smith, 2013; Holldober and Wilson, 2008). Nonetheless, either interpretation is compatible with an explanation in terms of the SR-Teleosemantic framework.

years of intense research, we are still discussing anuran representational mechanisms. In what follows, I would like to uncover some of the reasons that have led us to this situation.

2.1 Cognitive Representations

Can the SR-Teleosemantic framework sketched above be used to naturalize cognitive representations? Those who have tried to answer this question have tended to be rather pessimistic (Cao, 2012; Godfrey-Smith, 2013, 2013; Sprevak, 2013; cf. Ryder, 2006). This is, for instance, what Godfrey-Smith (2013, p. 17) claims:

> Senders and receivers are found in situations where there is *communication,* and the semantic phenomena involved in thought seem to be of a different kind. (...) One response to this is to say that surely the brain can be seen as a signaling device. Neurotransmitters transmit signals between neurons, for a start. But whether this kind of activity fits into the sender-receiver configuration discussed above is not so clear.

As Godfrey-Smith and others have suggested, the idea that some interactions between neurons instantiate the functional sender-receiver structure is suggestive. Nonetheless, making this idea precise turns out to be much harder than it looks. Many of the problems derive from key dissimilarities between explicit communication between organisms and internal cognitive signaling. The goal of this section is to identify some of these difficulties. I will first discuss some problems pointed out in the literature and I will argue that they are less troubling than might seem at first glance. Afterwards, I will present what I take to be the most pressing worries.

2.1.1 Some alleged difficulties

Just below the previous quote, Godfrey-Smith suggests a difficulty in using SR-TEL to naturalize cognitive representations:

> If we look inside the brain and find a huge network of neurons, each affected by some and affecting others, it appears that any neuron's firing might be described as either a signal, or the reception of a signal by the receiver, or the sending of another signal, depending of how one divides things up. This is not how things are in clear cases of sender-receiver configurations (2013, p. 17; see also Cao, 2012)

This passage (and similar ones in Godfrey-Smith, 2013 and Cao, 2012, 2014) can be interpreted in at least three different ways. First of all, one could interpret them as suggesting that there is no principled way of identifying the *mechanisms* that should play the role of senders and receivers. On this reading, the problem is that any classification of neurons and neuronal structures into significant units would be arbitrary. This claim, however, is probably too harsh. The task of different areas of cognitive science, such as neuroanatomy and neurophysiology, is precisely to localize functional components, describe them and investigate their patterns of interaction (Bechtel, 2008). In fact, there is a growing body of literature suggesting that most work in cognitive science is precisely devoted to the identification of mechanisms (Machamer et al. 2000; Glennan, 2002; Craver, 2007). Thus, unless some independent motivations are provided, we lack compelling reasons for doubting that the mechanisms that play the role of senders or receivers could be delimited without arbitrariness.

Alternatively, one might not be concerned with the identification of mechanisms, but with the claim that they function *as* senders or receivers. For instance, even if mechanisms A, B and C could be clearly discerned the worry would be that there is no principled way of establishing whether the receiver is A, B, A+B, A+B+C or many others. I do agree that in many cases this might a sensible worry (and a full answer would require a case-by-case analysis), but I doubt this difficulty might result in a general problem for a teleosemantic theory of cognitive representations. In the same way

that cognitive scientists distinguish mechanisms, they also make consistent hypotheses about their organization. Consider the recent debate on whether there is a dedicated system for face perception. The standard view claims that the three main areas primarily involved in face detection are the Fusiform Face Area (FFA), the Occipital Face Area (OFA) and the posterior Superior Temporal Sulcus (pSTS). These regions are widely thought to constitute the 'Core System' for visual analysis of faces, but their connection and the role that they play is still controversial (Haxby and Gobbini, 2010). According to one hypothesis, the OFA constitutes an early stage of face processing, which simply contributes to face detection. On this view, the OFA sends its output to the FFA, where face identification takes place (Haxby et al., 2002; Fairhall and Ishai, 2007; Ishai, 2008; Kanwisher and Barton 2010). However, according to an alternative view the initial coarse categorization of face stimulus takes place in the FFA, and then high-resolution details are analyzed in the OFA. An observation that has been suggested against the first hypothesis is the existence of some patients with structural damage in the OFA, which nonetheless can categorize a stimulus as a face (Rossion et al. 2011). Although in this context the expressions 'sender' and 'receiver' are rarely used, what seems to be at stake in this discussion is which areas in the Core System play the sender and the receiver role. The first hypothesis claims that OFA is sending signals to FFA and the second proposal inverts these roles. Obviously, a full analysis of the reasons that can be provided in favor of one or the other approach is far from trivial, but the idea that one cannot decide which mechanisms work as senders or receivers seems to be unfounded.

One could press this objection further and ask why the receiver is a particular region in the fusiform gyrus, rather than a larger area or even the whole temporal lobe (i.e. whether it is A, A+B or A+B+C). Interestingly, a similar question arose in this debate. Gauthier et al (1999) challenged the idea that there was a modular mechanism dedicated to face perception. They put forward the alternative hypothesis that activity in this area actually results from extensive experience, and that it will respond to any stimulus for which the subject has gained substantial perceptual expertise. To test this hypothesis, they trained subjects in the identification of novel objects that they called 'Greebles', and showed that after some experience identifying them the preference for faces over greebles in the fusiform gyrus of the right hemisphere was not significant. The reply by those who defend the modularity hypothesis required clarifying the receiver; they argued that adjacent to the Fusiform Face Area there is the Fusiform Body Area (FBA), which is dedicated to the perception of bodies (Kanwisher et al., 2006, p. 2113; Peelen and Downing, 2005; Schwarzlose et al., 2005). Thus, Gauthier and colleagues were considering a receiver that was too large, and that might explain why it failed to preferentially respond to faces. Again, in our terminology, their reply can understood as refinement of the receiver. In that particular case, the distinction between FFA and FBA was primarily based on a different pattern of stimulus response, but other strategies for distinguishing receivers could also be used. Therefore, the idea that there might be no principled way of establishing whether the receiver is A, B , A+B or A+B+C is unmotivated.

Finally, the previous quote can be understood as suggesting that, given the relatively weak definition of representational system endorsed by SR-Teleosemanticists, there would just be too many signals in the brain. One is likely to find sender-receiver structures at various levels and involving multiple structures. Note, however, that this would constitute a problem only if there was mismatch between a large set of signals implied by SR-Teleosemantics and a relatively sparse set of representations according to cognitive science. Yet this is an unlikely result. As a matter of fact, philosophers as well as scientists postulate an extremely large set of representations, which include action potentials in single neurons, transmitter releases, activity in organized populations of neurons and neural oscillations, among others (Bechtel, 2014; DeCharms and Zador, 2000, p. 624; Eliasmith, 2000, p. 64-5; Kandel et al. 2000, p. 30; Purve et al., 2008). The representational content of these signals can also diverge in many aspects; it can be propositional or non-propositional, represent external or internal states of the organism, stand for probability distributions, and so on (Gunther, 2003; Shea, 2014). Consequently, although a systematic mismatch between the signals

attributed by cognitive science and teleosemantics is an open possibility, there is no reason for thinking this will be the case.[4]

Therefore, although I agree that identifying mechanisms, classifying them into senders and receivers and specifying their signals is an arduous task, it is unclear that any of these features poses any insurmontable difficulty for teleosemantics. At the same time, however, I think these worries are backed up by a powerful intuition. Some aspects related to the complexity of cognitive systems seem to actually reveal important limitations of current teleosemantic models. Let me present some of them.

2.1.2 Some problems for a teleosemantic theory of cognitive representations

In this section I will present three specific difficulties for a teleosemantic theory of cognitive representations (connective complexity, isolation and feedback loops) and I will identify the general concern they probably derive from.

Connectivity

The brain is an extremely complex structure, in which many potential senders are actually connected to a large set of receivers. Nothing like a simple sender-receiver structure involving a single sender and receiver seems to be the case. The mapping between neurons or between neuronal networks is not one-to-one but many-to-many. Neurons in the central nervous system are usually innervated by thousands of synapses (although the number of synaptic inputs received by a neuron can vary from 1 to 100.000)(Purves, 2008, p.114-15). This complexity is hard to accommodate within the simple structure described in SR-TEL. Thus, the worry concerns the overwhelming *connectivity* of cognitive systems. Prima facie, it is not easy to see how the simple SR-Teleosemantics framework, which involves a single sender linked to a unique receiver in a direct way can be extended to account for the brain's labyrinthine organization.

Isolation

There is a second challenging feature of complex cognitive structures suggested by Cao (2012, p. 68):

> Most patterns of activation in sensory areas, especially in mammals like ourselves, are insufficient to cause a motor output. So these groups of neurons have the same problem of limited competence as the single neuron does. They can change the the brain state (now at a less local level, mobilizing other brain regions, for example), but they are still cut off from direct action in the outside world, being insulated from it by receivers lying between them and the muscles.(...) Only when our groups of neurons are organized such that they can close the loop from the external inputs to the external outputs can they be said to be fully competent actors on the world, rather than merely actors on the brain or representers of other brain regions.

Whereas in the simple sender-receiver model presented earlier the receiver´s output consists in a particular action being taken with respect to a certain environmental circumstance (e.g. meerkats responding to predators), this direct input-signal-behavior causal loop is lost in complex cognitive systems. As Cao (2012, p. 50) suggests, 'the world of the neuron (i.e. the world in which it is competent to take action) consists entirely of more neurons and the supporting cells around them'. In a nutshell, the worry is that most cognitive representations are *isolated* from the original input as well as from potential behaviors. Most cognitive mechanisms are triggered by other internal

---

4 Indeed, as a reviewer pointed out, even if there turns out to be more teleosemantic representations than cognitive science representations, it might still be possible to map one onto the other.

mechanisms and in turn lead to further activity downstream. It is difficult to see how these mechanisms, which are mostly surrounded by other brain structures and are usually situated many steps away from the environment, can qualify as senders and receivers according to SR-Teleosemantics. [5]

Feedback Loops

Finally, the simple sender-receiver structure seems to assume a strict hierarchical organization of representational mechanisms. It presupposes that information travels in a single direction from sender to receiver. However, a significant body of evidence that has been accumulating over the years show that this scenario is highly implausible. It is well-known, for example, that perceptual systems involve feedback as well as feedforward connections between representations. Many areas in early vision, for instance, receive information from higher-order visual areas (Goldstein, 2013, p. 24) or areas in the same hierarchical level (Lamme and Roelfsema, 2000). It has been suggested that these connections modulate the response to orientation (Ringach et al., 1997), color constancy (Wachtler et al., 1999), perceived brightness (Rossi et al., 1996) or perceptual grouping (Kapadia et al., 1995), among others. It is not obvious how feedback connections can be accommodated by the unidirectional and simple sender-receiver model suggested earlier.

How to expand teleosemantics?

The existence of multiple connections, relatively isolated representations and feedback loops constitute specific difficulties for an application of SR-TEL in the context of cognitive representations. This list is not supposed to be exhaustive, so similar difficulties could probably be added.[6]

Now, at this point one might be tempted to argue that these are just minor problems, which merely call for an extension and slight modification of the original sender-receiver model. For instance, one might appeal to additional tools, such as adapted functions, derived functions, stablizing functions and so on (Millikan, 1984). However, although some of these notions will of course be useful, it is unclear that they can fully address these worries. The fact that after three decades of research on naturalistic theories of representations these general questions still await a satisfactory answer should impel us to look for a deeper origin. As I pointed out, I think they probably stem from a more fundamental issue: we currently lack a general strategy for accommodating structures with a higher level of complexity. How can we extend the basic teleosemantic framework to accommodate more sophisticated structures? How can it be used to naturalize representational systems with intricate connections between multiple mechanisms, for instance? Can these structures evolve and acquire functions, in the sense required by teleosemantics? How can we study the representational properties of signals in these networks? Can we establish which of these states are representations and specify their representational contents? I think that, before the specific challenges posed by cognitive systems can be addressed, we should have a satisfactory answer to these questions.

---

5 In contrast, the fact that a close connection between stimuli and behavior obtains in fixed-action patterns (such as the anuran tongue-snapping mechanism) might explain why teleosemantics has been easily applied to these structures. Cao (2012) suggests that in humans this direct connection between input, signal and behavior is recovered if we focus on the whole cognitive systems and, as a result, she argues that only at this level can we properly talk of cognitive representations.

6 Cao (2012) puts forward additional worries, but most of them seem to derive from her own version of teleosemantics. According to her approach, representational status and content exclusively depend on the existence of an adequate receiver (so functional senders are not required). Similarly, she assumes that receivers can only be agents, in the sense that a receiver 'must be able to act in the world to affect its own outcomes' (Cao, 2012, p.53) and maintains that representational mechanisms need to be *flexible* and *competent*. As she convincingly argues, these features might generate additional difficulties for a teleosemantic theory of cognitive representation, but note that these issues just do not arise in the SR-Teleosemantic framework assumed here (which I think captures the central intuitions of some philosophers, such as Millikan, Shea or Papineau).

But how should we proceed? Initially, teleosemantics was probably inspired by communicative interactions between animals (including, of course, human language, see Millikan 1984). This motivation was very important for the sender-receiver model to take off, and it has certainly been extremely useful in modeling certain kinds of representational signals. However, for the theory to move forward, we might need to consider other models. This is precisely what I will try to do in the remainder.

3 Teleosemantic Models

In what follows, I will argue that the basic teleosemantic framework can be extended so as to overcome the previous limitations. The strategy I will pursue here consists of two steps. First of all, in this section I will show that teleosemantics can be described with the tools of evolutionary game theory. More precisely, some evolutionary models of signaling can be interpreted as providing the required tools for addressing systems that satisfy the conditions stated in SR-Teleosemantics. In the next section, I will use the instruments provided by models in evolutionary game theory to argue that the shortcomings of previous teleosemantic accounts can be avoided. Thus, if this strategy succeeds, it will vindicate the use of SR-Teleosemantics in the naturalization of cognitive representations.

3.1 Evolutionary Models of Signaling

The first step is to argue that some models in evolutionary game theory can be used to provide a more precise description of teleosemantics.[7] Thus, we first need to briefly present these models.

The models of signaling I would like to concentrate on derive from Skyrms (1996). Since then, there has been a growing body of literature in the area, which has developed this central idea in new and unexpected directions (e.g. Bruner et al. Forthcoming; Huttegger, 2007,2007; Huttegger and Zollman, 2011; Barrett, 2013; O´Connor, 2014; Skyrms, 2010, 2014). To illustrate this set of models, let us focus on the most basic example. The simplest model involves two players, A and B, which might be called 'sender' and 'receiver'. Suppose there are only two possible states of the world, S_1 and S_2, which only the sender can observe, and two possible actions, A_1 and A_2, which only the receiver can perform. Furthermore, let us assume that the sender has two messages available, M_1 and M_2. We call them 'messages' or 'signals', but it is worth stressing that (before evolution comes in — see below) they are just two meaningless states that the sender can produce and which can be detected by the receiver. Two further notions that need to be introduced are *strategy* and *common interest*. A sender's strategy specifies the set of messages that the sender would produce in every state of the world. More formally, it is a mapping from states of the world onto signals, i.e $F_S:S \to M$. Likewise, a receiver's strategy specifies the set of actions that she would undertake given a certain message. In other words, it is a mapping of signals onto actions, i.e. $F_r:M \to A$. Finally, this basic model supposes that both sender and receiver have common interest, in the sense that both get a positive payoff only if an action is performed in a specific state of the world. In particular, if A_n is performed in S_n both get a payoff of 1; otherwise, both get a payoff of 0.

It should be obvious that in this simple game both players face a coordination problem. Both are interested in a particular action being taken in a specific circumstance, but only the sender can

---

7 Interestingly, both teleosemantics and evolutionary models of signaling derive from Lewis' (1969) original work. Furthermore, they both try to dispense with the requirement of full rationality and knowledge assumed by Lewis (and classical game-theoretic approaches) in order to apply this framework to cognitively unsophisticated organisms. Nonetheless, despite this common origin, research in evolutionary game theory and the teleosemantic tradition have evolved independently and without much contact (but see, for instance, Harms, 2004 and Godfrey-Smith, 2013, 2013). One of the goals of this paper is precisely to call for a greater interaction between them.

observe the state of the world and only the receiver can act on it. If they interact just once (and assuming that they lack any previous convention or rational capacities), many results are possible. Senders and receiver have to choose one among different strategies; some give them a payoff of 1 and others give them a sub-optimal payoff. Nonetheless, in one-shot interactions they cannot know which strategy is the best, because they ignore the strategy followed by the other player. Among all possible sets of strategies, Lewis (1969) called a 'signaling system' a combination of sender and receiver strategies that guarantee that both players get the maximum payoff regardless of which state of the world occurs. For instance, the strategies $F_s = \{ <S_1, M_1>, <S_2, M_2> \}$ and $F_r = \{ <M_1, A_1>, <M_2, A_2> \}$ constitute a signaling system.

Some models in evolutionary game theory investigate how signaling systems evolve and eventually stabilize. These models usually assume a population of interacting individuals in which different strategies are played and examine the frequency changes through time in a population. Consider a simple example of one-population model. Suppose there is a large population in which individuals sometimes play the role of senders and sometimes pay the role of receivers. Thus, each individual has a strategy as a sender and as a receiver. For instance, as a sender, an individual might send M_1 in S_1 and M_2 in S_2 and, as a receiver, she might always perform A_1. We also suppose that individuals cannot change their strategies as senders and receivers. Furthermore, we make sure that at the beginning of the game all possible strategies are played by some members. Individuals interact randomly and, crucially, an individual's offspring depends on the payoff it gets. In other words, individuals that play more successful strategies have more children and become more frequent in the population in the next generation.[8] It has been shown that, in the simple game sketched above involving two equiprobable states, two messages and two actions, evolutionary dynamics will lead almost any initial population to a signaling system (e.g. Skyrms, 1996; Huttegger, 2007). Technically speaking, signaling systems are the only strict Nash equilibria in this game, which means that once this point is reached, any individual which unilaterally deviates does worse. All equilibria are unstable except for signaling systems. Although this result is quite intuitive and could have been predicted from the structure of the game, the result is robust and important: evolutionary dynamics shows that (under certain conditions- see Huttegger, 2007) in any population in which individuals play different strategies and interact randomly, signaling systems will emerge (Skyrms, 1996, 2010).

Now, why is all that relevant for our original question concerning teleosemantics? What is the relation between these models in evolutionary game theory and representations? Consider again the simple model sketched above. There is an initial population of sender and receivers, which follow different strategies. Individuals interact randomly with each other and the most successful strategies get a larger representation in the next generation. Thus, it could be said that strategies are *selected for* in a process that mimics natural selection. Accordingly, given the etiological definition of function provided earlier, that means that individuals[9] with more successful strategies acquire the *function* to act in certain ways given certain states of the world. More precisely, senders acquire the function of sending signals M_n when states S_n obtain and receivers acquire the function of performing actions A_n when signals are sent.

---

8 More precisely, these models usually assume replicator dynamics, which states that populations change according to the equation $\dot{x}_i = x_i(f_i(x) - \Phi(x))$, where x_i is the proportion of individuals of type i in the population, $x = (x_1, x_2, ...x_n)$ is the vector of the distribution of types in the population, $f_i(x)$ is the fitness of type i, and $\Phi(x)$ is the average population fitness.

9 Individuals in these models need not correspond to what we usually call 'individuals' or 'organisms'. They are just items that have certain effects and can be selected for. Thus, they can be interpreted as the devices that we have been calling 'senders' and 'receivers'.

As a result, models of signaling in evolutionary game theory provide a useful way of formulating and studying teleosemantics. After many generations, senders have been selected for producing certain signals in specific states of affairs, and receivers have been selected for performing certain actions when determined signals are sent. Thus, the intermediate signals qualify as representations. As a first approximation, then, within this paradigm SR-Teleosemantics can be formulated as follows: a state is a representation iff it is an intermediate state within an evolved sender-receiver structure described by these models.[10] Since the assumptions of some of these models partly overlap with the conditions that must be met for a representational system to exist according to SR-T$_{EL}$, signaling systems in these models seem to correspond to representational systems in the teleosemantic framework. Thus, they might offer a precise tool for stating the key ideas of teleosemantics and investigating central questions around representational systems: how signaling systems evolve, under what conditions is communication made easier or more difficult, what is the representational content of signals in complex networks or what is the relation between information and semantic content, among others.

Indeed, a teleosemantic interpretation is not restricted to the simplest games. Think, for instance, about models involving partial conflict of interest. Consider a game with three equiprobable states (S_1, S_2, S_2), three signals (M_1, M_2, M_3), three acts (A_1, A_2, A_3), and the payoff table in figure 1 (in each cell, the first number indicates the sender's payoff and the second the receiver's). In this scenario, both sender and receiver prefer A_2 to be performed in S_3 but disagree on the act to be undertaken in the rest of cases: whereas the sender prefers A_3 in S_1 and S_2, the receiver prefers A_1 in the first case and A_2 in the second. In other words, there is partial conflict of interest. An equilibrium in this game is a situation in which the sender sends signal M_1 in S_1 and S_2 and signal M_2 in S_3, while the receiver performs A_3 in receiving M_1 or M_3 and A_2 in M_3. Thus, a consequence of the lack of complete common interest is that the sender uses the same signal in two different states and the receiver performs the same action given two different signals (technically, this is called 'partial pooling'). This picture also seems to fit the SR-T$_{EL}$: the sender has the function of sending the same message (M_1) in two states (S_1, S_2) and the receiver the function of performing the same action (A_3) when these messages are sent. Thus, these models can also be used to develop teleosemantics in novel directions.[11]

Figure 1: *Payoff table with partial conflict of interest (Skyrms, 2010, p. 81)*

	A_1	A_2	A_3
S_1	2,10	0,0	10,8
S_2	0,0	2, 10	10,8
S_3	0,0	10,10	0,0

It is worth stressing, however, that despite the close connection between these models and teleosemantics, this link has largely been overlooked in the literature (but see Harms, 2004, 2007; Godfery-Smith, 2013). One of the reasons that might explain this neglect is the large variety and complexity and evolutionary models and teleosemantic theories. For one thing, teleosemantics comes in different versions; some of them stress the importance of the sender (Neander, 2013;

---

10 Again, this is receipe for state-types to qualify as representations. State-tokens will count as representational in virtue of belonging to this type.

11 A related and more difficult question is whether signaling can also evolve in situations of complete conflict of interest between sender and receiver, i.e. when the preference ordering over acts in every state is reversed. This is a disputed issue in both evolutionary game theory and teleosemantics (Godfrey-Smith and Martinez, 2013; Stegmann, 2009; Artiga, 2014) and, consequently, does not jeopardize the idea of using models of signaling as a teleosemantic tool.

Schulte, 2012) whereas others emphasize receivers (Cao, 2012, Jablonka, 2002). Similarly, some reject the cooperation requirement (Stegmann, 2009), others supplement teleosemantics with an additional informational condition (Shea, 2007; Martinez, 2013) and most of them disagree on how representational content is determined (Agar, 1993; Millikan, 1991; Neander, 1995; Price, 2001). In a parallel fashion, in evolutionary game theory there are models of signaling in which many of these parameters of the basic model presented here are altered (see below for more examples). As a result of this neglect, people working with evolutionary models usually reject teleosemantic approaches to representational content (Skyrms, 2010; Birch, Forthcoming). Skyrms (2010, p. 43-44), for instance, recently wrote:

> I say about teleosemantic intentionality the same thing I said about mentalistic intentionality. If we have a good model where it applies, it can be added to the theory. But neither intentionality nor teleosemantics is required to give an adequate account of the informational content of signals. Here I stand with Dretske [1981].

The idea I have trying to defend in this section is that teleosemantic intentionality is not an addition to the theory, but a different interpretation of the same models. They can be read as approximately providing the conditions stated in SR-Teleosemantics. For this reason, models of signaling can be used to investigate the conditions that must be met for an entity to represent another entity according to teleosemantics.[12] This perspective suggests an intriguing and I think fruitful research project: to investigate the links between the classical debate on naturalistic theories of content and mathematical models of signaling. This line of research might open a whole range of questions, whose answers might help each of these areas to move forward.

Of course, for obvious reasons in this paper I cannot attempt to fully address this issue. Nevertheless, since our goal is to use these models in the context of cognition, there is an important stumbling block that I would like to discuss: the nature of payoffs.

3.1.1 Payoffs

While there might be some general disanalogies between evolutionary models and teleosemantic theories, in trying to apply mathematical models in the naturalization of *cognitive* systems an additional difficulty arises: how to interpret of the notion of *payoff*. In classical game-theoretic analysis, payoffs usually represent the motivation of the player, such as the profit, quantity, utility or simply the desirability of a certain outcome. But, of course, if game-theoretic resources need to be used in the naturalization of representations, payoffs cannot be understood in that way. It is dubious, for instance, that neurons (or neuronal networks) have motivations or desired outcomes. So how should we interpret payoffs in this context and how can we measure them?[13]

One of the key aspects in which evolutionary game theory differs from classical game theory is in not presupposing that players are rational beings with complex intentional states, such as intentions, beliefs or desires. In these models payoffs are usually interpreted as a proxy for fitness. A payoff is a measure of the mean number of descendants left in the future that plays a given strategy (McNamara and Weissing, 2010). Furthermore, note that payoffs are not assigned to individuals, but to strategies (or to individuals individuated by their strategies), so what is crucial is that the payoff determines the probability that this strategy is adopted in the future; whether it is adopted by the same individual at later times or by its offspring is usually not essential. Accordingly, to specify the payoffs in the context of neuronal structures, we should identify the mechanisms that account for the spread, maintenance or elimination of certain strategies.

---

12 This is not to deny, of course, that these models can also be used by informational theorists

13 I would like to thank Kim Sterelny for pressing on this issue.

In that respect, Cao (2012, p.63-4) distinguishes three kinds of rewards that neurons obtain and that can account for the probability of their strategies being adopted in the future. First of all, a neuron can get a local reward in terms of a contribution to the individual cell's survival. Sometimes, this contribution only depends on the neurons' own activity. For instance, she mentions that after an energy-intensive peak of activity, the fact that astrocytes increase their lactate levels to feed neurons might be considered a reward. Another example is the phenomenon called 'cell-competition', in which weaker cells are eliminated from a growing population and substituted by stronger cells (Amoyel and Bach, 2014). Similarly, since we saw that neurotransmitter release is a kind of signal, coordinated activity between the pre-synaptic and post-synaptic neuron can also give rise to rewards. Pairing pre-synaptic and post-synaptic activity causes long-term potentiation (LTP), which strengthens the connection between two neurons (the opposing phenomenon is Long-term Depression, in which the set of synapses is weakened due to uncoordinated activity). These mechanisms warrant an attribution of payoffs to strategies, and in this case they determine the probability that the same neuron will pursue the same strategy at later times.

Neurons can also obtain a second kind of reward when they participate in a collective (e.g. a population of neurons that is supposed to perform a certain task) that does well, and thus increments the survival of all its parts. For instance, some learning mechanisms (e.g. operant conditioning) might involve heritability, mutation and differential replication, so in this context talk of rewards might be fully accurate (Hull et al. 2001). Indeed, this idea has motivated the development of learning game theory, which partly overlaps with evolutionary game theory.

Finally, neurons can get a distal reward when the organism reproduces. By increasing the probability of survival and reproduction, a neuron also increases the probability of future generations carrying neurons of the same type. In this case, the payoff does not measure the probability that the very same individual plays a certain strategy in the future, but how many of their descendants will adopt it. As we saw, however, this is not essential in evolutionary models.[14]

Of course, much more could be said about each of processes and the associated payoffs. Nonetheless, I hope I have at least dispelled the worry that talking of rewards in neuronal structures is inappropriate. This brief analysis reveals a complex interplay of different reward systems at different levels, and shows that there is nothing mysterious or confused about it.

3.2 Naturalizing Cognitive Representations

It is time to go back to cognitive representations. In the first part of the essay, we identified a general problem for the application of the basic teleosemantic schema to cognitive representations: we just do not know how to extend the basic teleosemantic framework to accommodate more complex structures. Given what we said in the previous section, however, we now have a strategy for addressing this question. To develop the basic teleosemantic framework sketched earlier, we just need to look at certain models in evolutionary game theory. To illustrate in more detail the strategy I am suggesting, in this section I would like to show how different models that have recently been presented in these areas can help overcome the three specific problems identified earlier: the complex connectivity of neurons, its relative isolation and the existence of feedback loops.

---

14 Two features of this analysis are worth stressing. First, the fact that the neuron's reward depends on the actitivity of many other brain structures should not be regarded as a problem. In general, traits are highly dependent on many others. The function of the gallbladder, for instance, is to store bile, because this is the effect that explains why the ancestors that possessed this organ outperformed those who did not, but a gallbladder can only produce this effect when many other traits are in place and functioning properly (liver, small intestine,...). Secondly, note that different reward systems might come into conflict. A neuron that follows a certain stategy might obtain a local reward without benefiting the organism, and vice versa. As a result, there might be a trade-off between them.

3.2.1 Multiple senders and receivers

As we saw, in the context of cognitive representations many potential senders are actually connected to a large set of receivers. Can this complex configuration be accommodated by SR-Teleosemantics? The suggestion is to address this issue by looking into evolutionary models of signaling. And, fortunately, there are already some formal models analyzing how these complex structures can originate. Indeed, interactions between multiple players have been studied at length from a game-theoretic point of view (Dixit and Skeath, 2009, p.55). From an evolutionary perspective, Skyrms (2009, 2010) has recently presented some models involving two senders and one receiver (See figure 2). In these models, there are four possibles states of the world, S_1, S_2, S_3 and S_4, and four actions A_1, A_2, A_3 and A_4. All players have complete common interest (i.e. both get a payoff of 1 if A_n is performed in S_n), but each sender can only make a partial observation. Let us suppose that sender 1 can only observe whether $\{S_1, S_2\}$ or $\{S_3, S_4\}$ and sender 2 whether $\{S_1, S_3\}$ or $\{S_2, S_4\}$. Together, both observations determine a single state of the world, but not separately. This schema exemplifies a system with two senders and one receiver. Skyrms has presented some evidence suggesting that this structure with two senders and a single receiver will evolve into a signaling system. In fact, Skyrms (2009) argues signaling systems are the only strict equilibria, which means that once a stable strategy is reached, any unilateral deviation does worse. Consequently, this dynamic model can be used to show that structures with multiple senders and receivers can be accounted for within the SR-Teleosemantic framework.[15] More complex models with various mechanism sending and receiving signals are currently under research (e.g. Barrett, 2013).

Figure 2: *Multiple senders and a single receiver.*
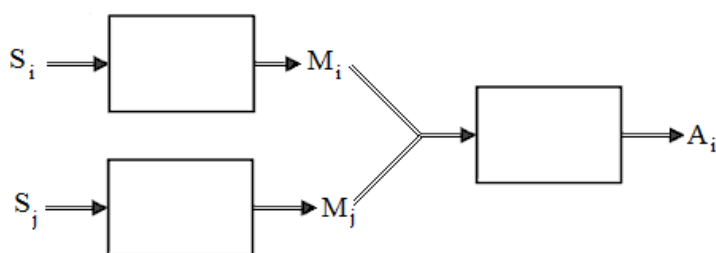
Not only are there formal models indicating that these structures can evolve. Real examples can be pointed out. For example, the neural circuity employed in sound localization consists of bipolar cells located at the Medial Superior Olive (MSO). Each bipolar neuron receives axons from both ears (more precisely, from the cochlea and cochlear nucleus). Yet, for obvious reasons, not all neurons are equally close the ears: some of them are closer to the right ear (so that the axon exiting the right cochlea is shorter than the axon exiting the left cochlea) and others are closer to the left ear. As a result, different neurons are maximally sensitive to different time delays, and for this reason they can be used as coincidence detectors. With this mechanism, the brain can locate the sound source only by combining the information from both ears; in the same way that in the signaling model the information provided by a single sender is insufficient, in sound localization it is not possible to locate horizontally a sound source from the signal arriving at single ear. For

---

15 A mechanism with two senders and one receiver is certainly a very simple structure so, as a reviewer suggested, one might question whether evolutionary game theory is required here. Nevertheless, two caveats are important to keep in mind. First, remember that the challenge posed by cognitive representations is that neurons are usually connected to hundreds or thousands of other neurons, and it is specially in this sort of cases that classical approaches face difficulties. Thus, although I focus on simple models for explanatory purposes, one should not forget that there are certain cases that traditional approaches have trouble accommodating (see above). Secondly, the goal of the paper is to argue that models in evolutionary game theory can be used to extend teleosemantics in new directions. This is fully compatible with there being some contexts in which other models or approaches are more illuminating.

bipolar cells to perform their function, the system has to employ the information provided by the two ears. Of course, there might be relevant aspects of the sound localization system that cannot be captured with this simple model, but nonetheless it seems to provide a promising approach. It is not unreasonable to suppose that these representations might be naturalizable by employing these or similar models.

3.2.2 Sequential Mechanisms

A different challenge is posed by the relative isolation of most representations from the corresponding environmental input and behavioral output. Most cognitive mechanisms are triggered by other mechanisms and in turn lead to further activity in other parts of the brain. It still has to be shown that the simple sender-receiver outlined earlier can accommodate a serial configuration.

Fortunately, these structures have also been widely studied in game theory. The most obvious instances are sequential games, in which there is a set of players and each of them makes a choice only after another player has already moved (Binmore, 2007, p. 37; Dixit and Skeath, 2009 p.55). Tic-Tac-Toe is a simple example of sequential game involving two players. These games are usually represented with a tree diagram, in which the root of the tree corresponds to the first move and the branches represent the choices that can be made at that move.

The evolution of multiple mechanisms linked sequentially has also been recently investigated. Suppose there are three players connected in a row and let us call them 'sender', 'intermediary' and 'receiver' (see figure 4↓). The sender observes one of two equiprobable states $\{S_1, S_2\}$ and chooses a signal from the set $\{B, C\}$. The intermediary observes the sender's output and sends a signal from the set $\{D, F\}$ to the receiver. Finally, the receiver observes the intermediary's signal and chooses an act from $\{A_1, A_2\}$. All of them get the same payoff conditional on the correspondence between the state of the world and the receiver's act (A_n when S_n holds). It has been shown that these structures can evolve into signaling systems (Skyrms, 2009, 2010). Moreover, they are the only strict equilibria. An example is provided by the following strategies: sender $\{<S_1, B>, <S_2, C>\}$, intermediary $\{<B, D> <C, F>\}$ and receiver $\{<D, A_1> <F, A_2>\}$.

Figure 3: *Sequence of sender and receivers.*

Note that, whereas in the original sender-receiver structure senders produce signals and receivers perform actions, in these more complex models involving a set of mechanisms in a row some players receive signals as input and deliver signals as output (in particular, this is the job of the intermediary). In the context of cognitive systems, this is probably the rule rather than the exception. A clear example is provided by mechanosensor receptors in tactile perception. There are three clearly differentiated stages in the path leading from the activation of skin receptors to the brain, which are usually labeled 'first-order', 'second-order' and 'third-order neurons'.[16] The axons of cutaneous mechanosensory afferents (first-order neurons) enter the spinal cord and synapse on neurons in the dorsal column nuclei (in the brainstem). Second-order neurons exiting the dorsal

---

16 'Higher-order' is used here in the psychological sense. Very roughly, in psychology this expression usually refers to states further downstream, while in philosophy it is employed for representational states that are about representational states.

column nuclei send their axons to the ventral posterior lateral nucleus of the thalamus. Finally, the third-order neurons send their axons to the primary and secondary somatosensory cortex (Purves et al., 2008, p. 218). Thus, instead of directly carrying information onto cortical areas, tactile information originated at the skin goes through two intermediaries. The formal model just presented seems to provide an illuminating representation of this neuronal organization.

3.2.3 Feedback

The last difficulty in applying the simple sender-receiver model to cognitive mechanisms is that connections often go in both directions. Cognitive processing involves feedback as well as feedforward links. Again, it is not obvious how feedback connections can be accommodated by the unidirectional and simple sender-receiver model suggested earlier, even if multiple or sequential mechanisms are added.

Indeed, models in which information flows in two directions have also been provided, although they are slightly more complicated (Skyrms, 2009). Suppose there are two players with common interest, player 1 and player 2. There are four possible states of the world (S_1, S_2, S_3, S_4) divided in two partitions ($P1:\{\{S_1, S_2\}, \{S_3, S_4\}\}, P2:\{\{S_1, S_3\}, \{S_2, S_4\}\}$). The observational partition of player 1 is not fixed, so that she can choose which observation to make. In turn, player 2 can face two different problems, which require choices in different action sets: $D1:\{A_1, A_2\}, D2:\{A_3, A_4\}$. Nature tosses a coin and decides whether player 2 is confronted with decision problem $D1$ or $D2$. Then, player 2 sends a signal to player 1, who observes the corresponding partition ($P1$ if the problem was $D1$ and $P2$ if the problem was $D2$). Depending on how the world turns out to be, player 1 sends a signal to player 2, who acts accordingly. The payoffs are given in figure 4. Once more, it has been shown that such interactions can evolve into signaling systems, which are strict equilibria.

Figure 4: A schematic representation of a dialogue situation and its payoffs.
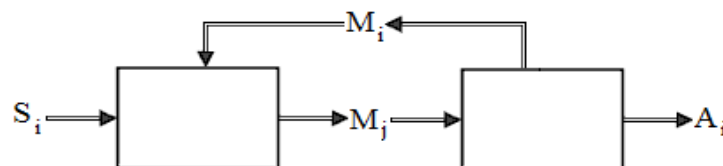
	Decision 1	Decision 1	Decision 2	Decision 2
	A_1	A_2	A_3	A_4
S_1	1	0	0	1
S_2	1	0	1	0
S_3	0	1	0	1
S_4	0	1	1	0

The primary visual cortex (V1) might provide an illustrative example of this configuration. The existence of feedback connections between V1 and higher cognitive areas is fairly well established, although their exact role is still under research (Salin and Bullier, 1995). Hupe et al. (1998) investigated the feedback connections between V1 and the middle temporal area called 'V5' (or 'MT') and found that a bar moving across the center of the visual field on a background of square checks tends to elicit a strong activity in certain areas of V1. To examine how activity in V5 and the surrounding area of the Superior Temporal Sulcus influence lower-order areas such as V1 or V3,

Hupe et al. ([1998](#)) inactivated these areas by cooling in macaque monkeys and presented them with a display of bars moving on a background of square checks. Their results showed that inactivity in V5 caused a significant diminution of responses in V1, V2 and V3, specially with low-salience stimuli. The effect it so strong that in certain cases V1, V2 and V3 remained completely silenced if feedback from V5 was blocked. They concluded that activity in V5 modulates the responses of V1 (and, specially, their contribution to figure-background segregation). Similarly, it has been suggested that covert attentional shifts can influence how V1 reacts to external cues. Gandhi et al ([1999](#)) gave subjects a discriminatory task in which they had to choose which of two circular stimuli presented at opposites sides of the visual field was spinning at a greater speed. Even though subjects were not allowed to move their eyes, when the relevant stimuli was presented on one side they found greater activity in the contralateral area of V1. They interpreted these results as suggesting that covert attentional shifts directed from higher-order areas were influencing how V1 responded to particular stimuli.

Obviously, these results have not been undisputed. Nonetheless, what is important for our purposes is that model presented earlier could accommodate them. In this case, V5 and attentional mechanisms correspond to player 2: given a certain task, they determine the observational partition for player 1, who modulates its response accordingly and informs player 2 (and other systems) about the state of the world. Thus, these models might help us to study and naturalize this kind of representations.

Summing up, the conclusion of this discussion on alternative configurations is that the basic SR-Teleosemantic framework can be extended to more complex structures and account for multiple sender and receivers, sequential representational mechanisms and top-down connections.[17] Consequently, evolutionary models of signaling provide a promising strategy for modeling and naturalizing a complex set of cognitive structures with the SR-Teleosemantic framework.

4 Conclusion

We started with a puzzle. Whereas the notion of representation is a central element in cognitive science, we still lack a general theory explaining what distinguishes representation from non-representations. The main goal of this essay was to argue that SR-Teleosemantics is a promising candidate for providing such a theory. More precisely, I identified some of the key limitations of the basic teleosemantic framework and argued that they can be overcome by using the powerful tools offered by evolutionary game theory. Of course, some difficulties are likely to remain. For instance, some cognitive structures might not fit the framework suggested here or might be too complicated for being modeled. Similarly, I have not addressed the question of what determines representational content. Nonetheless, I think that the approach defended here points at a wide range of tools for developing the classic teleosemantic theory in new and original ways. Indeed, if it is successful, this project would also show how various areas that employ the notion of representation (e.g. communication theory, genetics and cognitive science) can be unified and naturalized using the same paradigm. That achievement would constitute a significant step forward in our understanding of intentional phenomena.

In conclusion, I think the pessimism concerning the application of SR-Teleosemantics to complex cognitive mechanisms is probably unfounded. This idea suggests an ambitious research project:

---

17 I have presented these three types of models separately because they are supposed to address a different difficulty for SR-Teleosemantics, but there is a growing body of literature in game theory suggesting interesting connections between them (e.g. Bala and Goyal, [2000](#)). These more complex models will probably offer further tools for addressing cognitive phenomena

examining how far SR-Teleosemantics can be used in the naturalization of sophisticated representational states. This essay merely constitutes a first step in this direction.

Acknowledgements

References

N. Agar: "What Do Frogs Really Believe?", Australasian Journal of Philosophy, pp. 1-12, 1993.

C. Allen, M. Bekoff: Species of Mind: The Philosophy and Biology of Cognitive Ethology. MIT Press, 1999.

M. Amoyel; E. Bach: "Cell competition: how to eliminate your neighbours", Development, pp. 988-1000, 2014.

J. Barrett. The Evolution of Simple Rule-following. Biological Theory, 8(2):142-§150, 2013.

W. Bechtel. Mental mechanisms: Philosophical perspectives on cognitive neuroscience. Routledge, 2008.

W. Bechtel. Investigating neural representations: the tale of place cells. Synthese, 2014

K. Binmore. Game Theory: A Very Short Introduction (Very Short Introductions). Oxford University Press, 2007.

J. Birch. Propositional content in signalling systems. Philosophical Issues, Forthcoming.

S. Gandhi; D. Heeger; G. Boynton. Spatial attention aects brain activity in human primary visual cortex. Proceedings of the Natural Academy of Science, 96:3314-§3319, 1999.

J. Hupe; A. James; B. Payne; S. Lomber; P. Girard; J. Bullier. Cortical feedback improves discrimination between gure and background by v1, v2 and v3 neurons. Nature, 394:784-§787, 1998.

P. Salin J. Bullier. Corticocortical connectins in the visual system: Structure and function. Physiological Reviews, 75(1):107§-175, 1995.

B: Rossion; L. Dricot; R. Goebel; T. Busigny. Holistic face categorization in higher order visual areas of the normal and prosopagnosic brain: Toward a non-hierarchical view of face perception. Frontiers in Human Neuroscience, 173, 2011.

R. Cao. A teleosemantic approach to information in the brain. Biology and Philosophy, 27(1):49-§71, 2012.

R. Cao. Signaling in the Brain. Philosophy of Science, 81(5):891-901, 2012.

M. Manser; R. Seyfarth; D. Cheney. Suricate alarm calls signal predator class and urgency. Trends in Cognitive Sciencesy, 6(2):55§-57, 2002.

C. Craver. Explaining the Brain. Oxford University Press, 2007.

J. McNamara; L. White D. Purves; G. Augustine; D. Fitzpatrick; W. Hall; A. LaMantia. Neuroscience. Sinauer Associates, 2008.

C. DeCharms and A. Zador. Neural representation and the cortical code. Annual Review of Neuroscience, 23(1):613§-647, 2000.

A. Dixit; S. Skeath; D.Reiley. Games of Strategy. University of Chicago Press, 2009.

F. Dretske. Knowledge and the Flow of Information. The MIT Press, 1981.

C. Eliasmith. How neurons mean: A neurocomputational theory of representational content. Unpublished Dissertation, Washington University in St. Louis, 2000.

E. Kandel et al. Principles of Neural Science. McGraw-Hill, 2000.

J. Fodor. A Theory of Content and Other Essays. The MIT Press, 1990.

S. Glennan. Rethinking mechanistic explanation. Proceedings of the Philosophy of Science Association, 9:S342§-353, 2002.

J. Haxby; E. Homan; M. Gobbini. The distributed human neural system for face perception. Trends in Cognitive Sciences, 4(6):223§-233, 2002.

P. Godfrey-Smith. Complexity and the Function of Mind in Nature. Cambridge University Press, 1996.

P. Godfrey-Smith. On the theoretical role of "genetic coding". Philosophy of Science, 67(1):26§-44, 2000.

P. Godfrey-Smith. Signals, icons, and beliefs. In D. Ryder, J. Kingsbury, and K. Williford, editors, Millikan and Her Critics. Blackwell, 2013a.

P. Godfrey-Smith. Sender-receiver systems within and between organisms. Philosophy of Science, 2013b.

E. Goldstein. Sensation and Perception. Oxford University Press, 2013.

I. Gauthier; M. Tarr; A. Anderson; P. Skwudlarski; J. Gore. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nature Neuroscience, 2(62):568-§573, 1999.

V. Bala; S. Goyal. A noncooperative model of network formation. Econometrica, 68:1181-§1231, 2000.

Y. Gunther. Essays on Nonconceptual Content. MIT Press, 2003.

W. Harms. Primitive content, translation, and the emergence of meaning in animal communication. In D. Kimbrough Oller and U. Griebel, editors, Evolution of Communication Systems, pages 31§-48. Massachusetts Institute of Technology, 2004.

W. Harms. Information and Meaning in Evolutionary Processes. Cambridge University Press, 2007.

J. Haxby and M. Gobbini. Distributed neural systems for face perception. In J.Haxby A. Calder; G. Rhodes, M. Johnson, editor, Handbook of Face Perception. Oxford University Press, 2010.

J Bruner; C. OConnor; H. Rubin; S. Huttegger. David lewis in the lab: experimental results on the emergence of meaning. Synthese, Forthcoming.

D. Hull; R. Lanfman; and S. Glenn. A General Account of Selection: Biology, Immunology and Behavior. Behavioral and Brain Sciences. 24 (2), pp. 511-527, 2001

S. Huttegger. Evolution and the explanation of meaning. Philosophy of Science, 74(1):1§27, 2007a.

S. Huttegger. Evolutionary explanations of indicatives and imperatives. Erkenntnis, 66(3):409-§436, 2007b.

A. Ishai. Let's face it: it's a cortical network. NeuroImage, 40(2):415-§419, 2008.

S. Fairgall; A. Ishai. Eective connectivity within the distributed cortical network for face perception. Cerebral Cortex, 17(10):2400-§2406, 2007.

E. Jablonka. Information: its interpretation, its inheritance, and its sharing. Philosophy of Science, 69(4):578-§605, 2002.

N, Kanwisher and J. Barton. The functional architecture of the face system: integrating evidence from fmri and patient studies. In J.Haxby A. Calder; G. Rhodes, M. Johnson, editor, Handbook of Face Perception. Oxford University Press, 2010.

R. Schwarzlose; C. Baker; N. Kanwisher. Separate face and body selectivity on the fusiform gyrus. The Journal of Neuroscience, 25(47):11055§-11059, 2005.

C. Westheimer G. Kapadia M.; Ito M.; Gilbert. Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in v1 of alert monkeys. Neuron, 15(4):843-§56, 1995.

V. Lamme and P. Roelfsema. The distinct modes of vision oered by feedforward and recurrent

processing. Trends in Neuroscience, 32(11):571§-579, 2000.

D. Lewis. Convention: A Philosophical Study. John Wiley and Sons, 1969.

A. Rossi; Rittenhouse C.; Paradiso M. The representation of brightness in primary visual cortex. Science, 173:110-4§7, 1996.

M. Martinez. Teleosemantics and indeterminacy. Dialectica, 67(4):427§-453, 2013a.

P: Godfrey-Smith; M. Martinez. Communication and common interest. PLOS Computational Biology, 9(11), 2013b.

J. Maynard-Smith. The concept of information in biology. Philosophy of Science, 67(2):177-§194, 2000.

J. McNamara and F. Weissing. Evolutionary game theory. In T. Szekely; A. Moore; J. Komdeur, editor, Social Behavior. Genes, Ecology and Evolution. Cambridge University Press, 2010.

R. G. Millikan. Language, Thought and Other Biological Categories. The MIT Press, 1984.

R. G. Millikan. In Defense of Proper Functions. Philosophy of Science, 56(2): 288-§302, 1989.

R. G. Millikan. Speaking up for darwin. In B. Loewer and G. Rey, editors, Meaning in Mind. Fodor and his Critics, pages 151-§164. Blackwell, 1991.

K. Neander. Functions as Selected Eects: The Conceptual Analyst's Defence. Philosophy of Science, 58:168§184, 1991.

K. Neander. Misrepresenting & Malfunctioning. Philosophical Studies, 79:109§-141, 1995.

K. Neander. Toward an informational teleosemantics. In D. Ryder; J.Kingsbury;

K. Williford, editor, Millikan and her critics. Wiley-Blackwell, 2013.

C. O'Connor. Evolving perceptual categories. Philosophy of Science, 81:840§-851, 2014.

L. Darden P. Machamer and C. F. Craver. Thinking about mechanisms. Philosophy of Science, 67(1):1§25, 2000.

D. Papineau. Reality and Representation. Basil Blackwell, 1987.

M. Peelen and P. Downing. Selectivity for the human body in the fusiform gyrus. Journal of Neurophysiology, 93:603-§608, 2005.

C. Price. Functions in Mind. Oxford University Press, Oxford, 2001.

Ringach D.; Hawken M.; Shapley R. Dynamics of orientation tuning in macaque primary visual cortex. Proceedings of the Aristotelian Society, 387(6630):281-4, 1997.

D. Ryder. On thinking of kinds: a neuroscientic perspective. In G. MacDonald and D. Papineau, editors, Teleosemantics. Oxford University Press, 2006.

T. Albright; T. Wachtler; T. Sejnowski. Interactions between stimulus and background chromaticities in macaque primary visual cortex. Investigative Ophtalmology and Visual Science, 40, 1999.

N. Shea. Consumers Need Information: Supplementing Teleosemantics with an Input Condition. Philosophy and Phenomenological Research, 75(2):404-§435, 2007a.

N. Shea. Representation in the genome and in other inheritance systems. Biology and Philosophy, 22:313§-331, 2007b.

N. Shea. Neural signalling of probabilistic vectors. Philosophy of Science, 81 (5): 902-913, 2014.

P. Shulte. How frogs see the world: Putting millikan's teleosemantics to the test. Philosophia, 40:483§496, 2012.

B. Skyrms. Evolution of the Social Contract. Cambridge University Press, 1996.

B. Skyrms. Evolution of signalling systems with multiple senders and receivers. Philosophical transaction of the Royal Society of Biological Sciences, 364(1518):771§-779, 2009.

B. Skyrms. Signals: Evolution, Learning, and Information. Oxford University Press, Oxford, 2010.

B. Skyrms. Social Dynamics. Oxford University Press, Oxford, 2014.

M. Sprevak. Fictionalism about neural representations. The Monist, 96:539§-560, 2013.

D. Stampe. Towards a Causal Theory of Linguistic Representation. Midwest Studies in Philosophy, 2:42§63, 1977.