A Brief Introduction to the Guidance Theory of Representation

Gregg Rosenberg (ghrosenb@ai.uga.edu)
Artificial Intelligence Center, University of Georgia
Athens, GA 30602 USA

Michael L. Anderson (anderson@cs.umd.edu)
Institute for Advanced Computer Studies, University of Maryland
College Park, MD 20742 USA

Abstract

Recent trends in the philosophy of mind and cognitive science can be fruitfully characterized as part of the ongoing attempt to come to grips with the very idea of *homo sapiens*—an intelligent, evolved, biological agent—and its signature contribution is the emergence of a philosophical anthropology which, *contra* Descartes and his thinking thing, instead puts doing at the center of human being. Applying this agency-oriented line of thinking to the problem of representation, this paper introduces the guidance theory, according to which the content and intentionality of representations can be accounted for in terms of the way they provide guidance for action. We offer a brief account of the motivation for this theory, and a formal characterization.

Introduction and Background

Recent trends in the philosophy of mind and cognitive science can be fruitfully characterized as part of the ongoing attempt to come to grips with the very idea of *homo sapiens*—an intelligent, evolved, biological agent—and its signature contribution is the emergence of a philosophical anthropology which, *contra* Descartes and his thinking thing, instead puts doing at the center of human being. Work that falls under this broad umbrella includes accounts of human cognition which stress embodiment and environmental situatedness (Anderson, 2003; *forthcoming*-a; Ballard et al., 1997; Clancey, 1997; Clark, 1995; Varela, Thompson, & Rosch, 1991), pragmatic and evolutionary accounts of human knowledge and culture (Barkow, Cosmides, & Tooby, 1992; Guignon, 1983; Hacking, 1983; Munz, 1993; O’Donovan-Anderson, 1997; Rescher, 1990) and action-oriented accounts of perception (Aloimonos, 1992; Ballard, 1991; Gibson, 1966; Milner & Goodale, 1995; O’Regan & Noé, 2001), to name only a few categories, and a few of the many works in each. The current essay introduces the results of our effort to build a theory of representation—in terms of some other, presumably more basic relation. For instance, a typical causal theory of representation might hold that a given representation *R* is about *E* just in case it has a certain specified set of causal relations to *E*, for instance, that perceiving an instance of *E* will cause one to represent with *R* (Fodor, 1981; 1987). Likewise an information-content approach might hold that a given representation is about that object from which the information it contains in fact derived (Dretske, 1981; 1986; 1988). Conceptual role theories, on the other hand, try to analyze meaning in terms of the role played by the concept in inferential and other conceptual/cognitive processes: roughly speaking, the representation *R* is about *E* just in case it is used to make warranted inferences about *E* (Harman, 1982; 1987). Naturally, there are also theories that try to combine these two approaches, producing the so-called “two-factor” accounts (Block, 1986; Loar, 1981; Lycan, 1984). There is no need, nor is this the place, to rehearse the standard critiques of these various theories (but see Anderson, *forthcoming*-b). However, by way of situating and introducing our own account of representational content, let us say that we find the various causal approaches too *input focused*, meaning they give too much importance to the ways in which the environment affects the organism to endow its states with representational meaning, and while the conceptual role theories seem to us a step in the right direction in that they draw attention to the importance of cognitive actions taken by the subject with its representations, *none* of the theories outlined above give sufficient weight to the full range of what a subject *does* with its representations.

In contrast, we ask first not what a representation is, but what it *does* for the representing agent, and what the agent does with it; what is a representation for? Our contention is essentially that representations are what representations do, and that what a representation *does* is provide guidance for action. Whatever the details of its instantiation or structure, whatever its physical, informational, or inferential features (and these are quite various across different representing systems), what makes a given item *representational* is its role in providing guidance to the cognitive agent for taking actions with respect to the represented object. In our view, each of those other special features a given representing token might possess—e.g. co-variance with,
openness to the causal influence of, or resemblance to its object—correspond to one of the range of strategies that our various representation-forming and representation-consuming systems have evolved to solve the biologically fundamental problem of providing autonomous organisms with guidance for action.

On the guidance theory action is fundamentally intentional: it is first and last a directed engagement with the world. Our basic claim is that representations come into existence and derive their content from their role supporting the basic intentionality of action. The fact that subjects take action with respect to things is what confers content on representations; it is how representations reach outside the organism and touch things in the world. The guidance theory presumes, then, that the intentionality of representation can be grounded in the intentionality of action.

A Formal Account of the Guidance Theory

Let us say that a token provides guidance to a subject by making its features available to the subject’s motor systems and rational control processes for use in making discriminating choices between possible actions or possible ways of executing actions. Below we introduce the foundations of the guidance theory in terms of a set of propositions, which together characterize the most central features of the theory.

1. An entity is anything that can be represented: a property, a concrete particular, an aspect of a thing, a state of affairs, a number, etc.

2. A subject is any representation-consuming cognitive engine. To be a representation consumer, it must be capable of interacting in the world in a rational, goal directed way due at least partly to guidance it receives from tokens within its cognitive systems.

3. A circumstance is a circumstance of the subject. A circumstance consists in the subject’s internal states, including the subject’s bodily changes, registrations, representations, expectations, priorities, values, options for action, homeostatic self-evaluations, procedural knowledge, motor schemas and also the subject’s immediate environment.

4. A subject standardly uses tokens (of a type) to provide guidance with respect to an entity E in a given (type of) circumstance C if, and only if, the subject has an enduring conscious preference or condition reflex to use the tokens (i.e., members of the type) to provide guidance with respect to E when in circumstance C.

5. An action can be a motor process or a cognitive process. This yields two clauses in the definition of action:

5.1 In the case of a motor process, a motor process is an action if, and only if, it is activated under control of perceptual/cognitive feedback processes capable of effectively modulating or bringing about changes in the organism or in the world.

5.2 In the case of cognitive processes, a cognitive process is an action if, and only if, it is a mental process under intentional control whose results contribute to circumstances (as defined above) used to direct motor processes. A cognitive process is under intentional control if the working of that cognitive process is subject to modification by processes of attention, short-term memory, valuation, assent and dissent, practiced learning, and consciously administered self-criticism and praise.

As mentioned already, the fact that subjects take action with respect to things is what confers content on representations; it is how representations reach outside the organism and touch things in the world. The central importance of the intentionality of action means that it is vital to correctly understand—without regress—what it is for an action to be taken with respect to something.

6. An action is taken with respect to an entity E if, and only if,

6.1 The action is a motor program, E is the focus of the intended change or efforts at control in the world; or

6.2 The action is a motor program and an assumption of information about E is a motivating reason that the given action, rather than some alternative non-E involving action, was undertaken; or

6.3 The action is a cognitive process undertaken to discover or confirm facts, to modify values, or to decide between alternative actions, and an assumption of information about E is necessary if the process as a whole is to provide guidance for the subject’s motor actions.

This definition uses three further terms—motivating reason, focus, and assumption of information—that present the potential for regress and require further discussion.

Motivating Reason

For an account of motivating reason, we hold only that any analysis must be such that it would be applicable to goal-directed behavior of entities that do not have representations at all. For example, it must be of a piece with how we would identify the motivating reasons for why a plant turns toward the sunlight. The plant’s behavior is goal-directed behavior even if it is not action in the sense defined above, and the motivating reason for the behavior is to maximize the amount of sunlight available for photosynthesis. Because the plant does not have representations, a correct account of motivating reason cannot appeal to representational content.

We also distinguish motivating reasons from applications of causal force. A child may go to bed early on Christmas Eve to encourage Santa Claus to bring presents, and this may be the child’s motivating reason, even though Santa Claus is not capable of applying causal force on the child’s mind. A hungry wolf may look for prey and its motivating reason may be a future state of satiety, even if the cause of its behavior is a present internal state. Any account of motivating reasons must allow for motivating reasons that are non-representational facts and entities, even for agents that possess representations.
At its heart, the concept of a motivating reason is deeply tied to concepts of rational interpretation like the one found in Daniel Dennett’s description of the intentional stance (Dennett, 1987). We take no position here on the basis of, or constraints on, any specific standards of rational interpretation.

Focus
As it is used above, the idea of an action’s focus is intended to express a functionalist concept. When a subject is performing an action it places itself into a potential feedback loop with its environment. Its purpose is to monitor the result of the action and to plan adjustments to its course of action.

(7) The focus of an action is the ultimate entity being monitored through the feedback channels taken to provide indications of its status.

A subject may monitor the focus directly, or indirectly by monitoring the status of some entity being used as an indicator of facts about the focus. Because indicators are made part of an extended guidance control system, indications about the focus will cause in the subject beliefs, decisions or equivalent states about further appropriate actions or perhaps that action may cease. When the focus is monitored through an indicator the subject may have an indirect causal connection to the focus or even no causal connection at all. An example of an indirect causal connection to a focus would be an engineer monitoring a gauge that is itself monitoring engine pressure. Examples of foci to which there is no causal connection are things like the time of day or a mathematical operation on numbers. To monitor the first we might monitor an indicator like a clock face and to monitor the second we might monitor a progression of numerals manipulated according to established rules. In both of these cases the focus of the action is something that is not present and to which the subject is not even indirectly causally connected, but which can be monitored nevertheless, despite the lack of causal causation, by establishing a connection to something else that can be manipulated to vary systematically with facts about the focus.

Identifying the focus of an action in a given case requires establishing the facts about what the subject is monitoring in its circumstances, and understanding these facts in terms of the subject’s motivating reasons.

Assumption of Information
An assumption of information is to be cashed out in terms of facts about the actual operation of the representing system (or subject). Beginning with an example will make the concept easier to grasp. Imagine a computer processing a user’s command to print a document. To do this, the computer must determine to which printer it should send its own commands. To guide this action, the computer reads several character strings contained on its hard disk, one identifying the printer and others with other information about the printer. These strings guide it regarding where it should send its print commands and what protocol it should use to communicate with the printer. From the perspective of the guidance theory, here is the key fact: these character strings represent what they do both because of the circumstances in which the computer is reading them and also because of the assumptions built into those circumstances. The computer processes the strings as if they conveyed information about the printer to which it sends its commands and which communication protocol it should use. There is no regress involved in claiming it makes this assumption, because the assumption itself is not a matter of having representational content. There is no representation inside the computer with the content: I assume that this string has information about the printer. Even more strongly, its ability to make an assumption of information does not require that the computer actually possesses information, nor that it ever did. In the case described, the character string the computer accesses could have been placed on the disk via the output of a random number generator and by coincidence be effective in directing it to the printer. Even were that to be true, the string still would be providing guidance and the computer would still be making an assumption that the string contained information about the correct printer. Therefore, the ability to make an assumption of information does not require an ability to have or obtain information.

Rather, the assumption of information about the printer is a matter of know-how that is built into the architecture of the computer: how it accesses representations, in what circumstances it accesses them, how it reads and interprets their structure, what actions it initiates and monitors upon accessing them, how those actions cause it to interact with the world, and so forth. We can provide a candidate analysis of this know-how. To do this, we first need to define, for any given token, the class of actions it supports. The class of actions a token \( T \) supports is relative to the kinds of circumstances \( C \) where the system is prepared to use the token for guidance. It consists of all the actions the system can initiate or modulate in \( C \) due to its processing of \( T \). Let us label this class of supported actions \( A_{\text{supp}} \).

(8) An action \( A \) is a member of the class of actions, \( A_{\text{supp}} \), supported by a token \( T \) used by a subject \( S \) in circumstances \( C \) if, and only if, \( S \) in \( C \) would use \( T \) for guidance regarding the initiation or manner of execution of \( A \).

We should think of the actions in \( A_{\text{supp}} \) as focus-neutral descriptions of an action in need of association with a focus in particular initiations. So, for example, if in some circumstances a system is prepared to use a token for guidance in running, the action running is the focus neutral description. If the specific initiation of this action occurs when the focus of the action is a bear, the focus-neutral action “running” is initiated as the focus-specific action “running away from a bear.” Actions obtain a focus in the way discussed above.

Furthermore, since subjects do not initiate actions at random, for each action in \( A_{\text{supp}} \), there will be a (possibly

---

1 This, assuming that possessing information depends on causal history and connection, which may not be the case.
very large but) finite set of circumstances capable of triggering the initiation of the action. We can call this set of triggering circumstances $A_{\text{circ}}$. The number of triples $<A \in A_{\text{supp}}, C \in A_{\text{circ}}, \text{Focus}>$ representing supported actions $A$ initiated in circumstances $C$ with focus $\text{Focus}$ provides a class of counterfactual action scenarios, $A_{\text{scene}}$, in which the token $T$ provides guidance for a subject. These are the action scenarios in which $T$ participates.

Most actions are complex, both in the sense that they have many different specific features that must be managed (e.g., the trajectory and velocity of a running motion), and in the sense that they almost always require initiating smaller or tangential actions involving entities besides its focus if they are to succeed in affecting their intended change or control (e.g., jumping over the branch on the ground while running from the bear). Because of the complexity of action, subjects needing to execute an action will almost always use representations other than the tokens representing the focus of the action. In fact, activation of these further tokens is necessary to fill out the circumstances in which all the tokens are used.

These other active representations will fall into several categories: conscious representations with foci of their own serving the larger action program; unconscious but potentially conscious representations supporting the interpretation of the circumstances and manner in which the action is executed; and sub-conscious representations that can never be conscious but that provide support for basic perception, adjusting bodily movements, and triggering emotion. We should construe the entities towards which the supporting tokens provide guidance as sub-foci in sub-actions lying under the umbrella of the main action. Therefore, these further tokens, the ones that support the guidance for the main action within a given $A_{\text{scene}}$, have functional roles determined by their potential relationships to their own foci within the circumstances $C$ of $A_{\text{scene}}$.

Relative to these action scenarios, the guidance theory supposes that in each $A_{\text{scene}}$ where an active token succeeds in having reference the token can be mapped to an entity through its functional role under the rational constraints associated with assigning motivating reasons to their sub-actions. This supposition is justified because, in providing guidance, a token will make features of itself available to the subject, which the subject can use to differentially control its actions with respect to an entity which is a focus or sub-focus of a given action.

The know-how involved in an assumption of information, then, is a question of the way that the subject’s decoders and action mechanisms process and/or respond to representations (i.e., how it accesses representations, in what circumstances it accesses them, how it reads and interprets their structure, what actions it initiates and monitors upon accessing them, how those actions cause it to interact with the world, and so forth) given the subject’s capabilities, needs, environment, and cognitive architecture. The general idea is that assumptions of information consist in non-representational facts about how the subject works, not in further representational facts about, or representations used by, the subject. Although this account is clearly preliminary, it does at least show how the idea of an assumption of information can be interpreted, and used as part of the machinery involved in determining the content of a representation, without initiating a vicious regress or involving circular appeals to representational content.

This brings us, finally, to the cumulative definition of representation. On the guidance theory, representation is simply tracking in the sense defined below:

(9) A token $T$ tracks an entity $E$ for a subject $S$ in token circumstances $C$ if, and only if, $T$ is standardly used to provide guidance to $S$ for taking action with respect to $E$ in $C$.

(10) A token $T$ represents an entity $E$ for a subject $S$ in token circumstances $C$ if, and only if, $T$ tracks $E$ for $S$ in $C$.

By linking representation to guidance in this way, the guidance theory distributes responsibility for the existence of representational content across a representational token (the representation) and an interpretative decoding mechanism (the decoder) integrated with a subject’s action-determining processes. The effect of distributing responsibility is to introduce new degrees of freedom regarding the exact physical or informational requirements for something to be a representation, as the requirements on the representation will depend on the capabilities of the decoder and the circumstances in which it is used. In general, the demands on each part of the coupled system vary inversely with the demands on the other. A representation that is highly structured and closely coupled with what it represents needs a less sophisticated decoding mechanism, while a very sophisticated (or very rigid and simple) decoding mechanism may embody (or presume) so much implicit domain knowledge that it can get by with very sparse representations.

**Representation and Misrepresentation**

One of the most important problems that any theory of representation must solve is the problem of normativity: representations are assessable for accuracy, and therefore they can be in error. To be complete, the guidance theory must account for this feature of representations. Because the guidance theory is an action-based theory of representation, the natural thing to do is to base error on the failure of action and the way that a representation’s guidance contributes to that failure. The intuitive idea, then, is that a representation is in error if it provided guidance to an action that failed in its intent, and it failed partly or wholly because of the guidance provided by that representation. This intuitive idea can be formalized as follows:

(11) An action fails in its intent if, and only if,

$(11.1)$ It is a motor action and the intended change is not achieved or the intended process is not brought under control; or

---

2 The concept of error will be defined formally in the next section.
(11.2) It is a cognitive process and it (a) confirms a representation that is in error; or (b) disconfirms a representation that is not in error; or (c) modifies a value in a way that the subject later regrets; or (d) recommends a course of action that fails.

(12) An action A fails in its intent because of R if, and only if, (a) A failed; and (b) A was taken with respect to an entity E; and (c) R provided guidance for A w.r.t. E; and (d) R has feature F; and (e) R with F represents that E has property P; and (f) A failed because E was not P. Note that the term “represents” in clause (e) is to be read in light of the current theory of representation.

(13) A token representation R is in error for subject S and action A in token circumstances C if, and only if, A would fail because of R if taken by S in C.

The representation may be said to be in error for S simpliciter if and only if the class of actions for which R provides guidance in S’s circumstances C is dominated by actions that would fail because of R.

Comparison to Related Work

The guidance theory, broadly speaking, takes both a naturalistic and a functional perspective on representation. It is motivated by the same fundamental insight regarding the epistemic importance of action and interaction as gave rise to the theory of interactive representation (Bickhard, 1993; 1999). However, we offer a significantly different development and formalization of this shared insight. For instance, Bickhard’s analysis relies heavily on control theory, cashes out representational content in terms of ‘environmental interactive properties’, and assumes some version of process ontology. The guidance theory, while compatible with these possibilities, does not require them. Still, the relative advantages of these two analyses remain largely to be determined. While there are many other naturalistic theories of representation on offer, very few adopt the functional perspective in as thoroughgoing a way as we do. For instance, Dretske (1986; 1988) adopts the functional perspective largely as a post-hoc fix to what remains an information-content approach to representation, so as to be better able to account for misinformation. In contrast, Ruth Millikan does take the functional perspective as the starting point for her theory of representation, and the guidance theory thus bears the most resemblance to hers (Millikan, 1984; 1993). Thus, although the current article is meant only as a concise introduction to the guidance theory, and is not the place for any detailed comparisons with rival theories, it is nevertheless worthwhile to say a few words about Millikan’s theory in particular.

The resemblance between the guidance theory and Millikan’s own biologically inspired theory is strongest when she writes things like: “Cognitive systems are designed by evolution to make abstract pictures of the organism’s environment and to be guided by these pictures in the production of appropriate actions.” (Millikan, 1993:11) However, the impression of similarity fades quickly as the details are examined. For while we agree on this very general characterization of cognitive systems, we differ as to the core point: that mental representations must be pictures and, even when they are pictures, we differ as to what makes such “abstract pictures” representations.

There are three main components to this very basic disagreement. First, on our view, a given mental token is a representation just in case it is standardly used by a given organism to guide its behavior with respect to the intended object; Millikan, in contrast, suggests that it is only a representation if it is the result of (or consumed by) a properly functioning system, performing the function it was selected to perform: “It is not the facts about how the system does operate that make it a representing system and determine what it represents. Rather, it is the facts about what it would be doing if it were operating according to biological norms.” (Millikan, 1993:10-11)

Second, and deeply related to the first, Millikan relies heavily on the notion of such a “proper function” to explain the possibility of representational error (a representation is in error when the relevant representation-producing or representation-consuming system is not functioning according to biological norms). In contrast, our theory allows for the possibility that a system serving some function other than that for which it was selected, or mal-functioning in some very lucky way, could, in its use of mental tokens, be representing just in case (roughly speaking) the mental tokens in question were being used to (successfully) guide the agent’s actions with respect to the indicated objects. Rather than analyze representational error in terms of mal- or non-standardly-functioning systems, we cash it out in terms of failure of action. Although we think representational systems did evolve, and attention to their evolutionary history can help us understand how and why they function as they do, we believe a system can sometimes competently perform a function, including representing, for which it was not selected, and in these cases its unusual provenance should be no barrier to recognizing this fact.

Third (and finally), whereas Millikan’s view of behavior and action revolves around the function or purpose of the organism or its parts (a movement by the organism is only a behavior of that organism if it can (or perhaps must) be understood in terms of the organism’s proper function or biological purposes), our own definition of action includes motor and cognitive processes effected for a broader range of motivating reasons. Although some element of teleology is apparently necessary to ground the idea of a motivating reason for acting, it is not clear to us that this must necessarily be accounted for in terms of natural selection. It could be the teleology of the subject itself, understood as having a subjective purpose like maintaining its homeostatic condition, pursuing hedonic value, or maintaining adherence to a moral, political, or aesthetic principle. A more detailed discussion relating the guidance theory to some alternative theories, including Millikan’s, can be found in (Anderson, forthcoming-b).
Conclusion

The guidance theory is an action-focused theory of representation according to which content is derived from the role a representational vehicle plays in guiding a subject’s actions with respect to other things. What qualifies an element of experience as a representation is, strictly speaking, only that the element of experience be capable of providing a subject with guidance for its actions with respect to entities. To be capable of providing guidance an element of experience only needs to have features useful for exploitation by the subject’s action-producing mechanisms.

In the full formalization, we show that the guidance theory can account for various problem cases of representational content such as abstract, fictional and non-existent objects (Rosenberg & Anderson, forthcoming). Twin-Earth and swampman are discussed in (Anderson, forthcoming-b). Future work will consider the evolutionary development of representation in more detail, and the implications of the guidance theory for the correspondence theory of truth, for scientific realism, and for consciousness and phenomenal content (Rosenberg, 2004).

References