Robotics and Autonomous Systems

Book review

Evans' *Varieties of Reference* and the anchoring problem

Michael L. Anderson

Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA

To think about how to anchor abstract symbols to objects in the world is to become part of a tradition in philosophy with a long history, and an especially rich recent past. It is to ask, with Wittgenstein, "What makes my thought about him, a thought about *him*?" and thus it is to wonder not just about the nature of referring expressions or singular terms, but about the nature of referring beings. With this in mind I hereby endeavor—briefly, incompletely, but hopefully still usefully—to introduce what in my judgment is the single best philosophical starting-point for those interested in understanding the referential connections between symbols and the world, and the cognitive, epistemic, and linguistic capacities which support them: *The Varieties of Reference* by Gareth Evans.[1]

It is worthwhile first of all to note, as the title indicates, that it is the *varieties* of reference that are of interest. It is Evans' contention that no single theory can account for our various use of singular terms; although the different kinds of reference share certain features, and rely on related cognitive, linguistic and epistemic capacities, it appears that, rather than being a class defined by necessary and sufficient criteria for membership, they form a family of abilities, united, like a thread, by its overlapping fibers.

Evans does not defend this claim so much as display it in his account. Much of the underlying variety in reference can be brought out by considering the guiding principle of the work as a whole, which Evans calls Russell's Principle: "The principle is that a subject cannot make a judgment about something, unless he knows which object the judgment is about" (p. 89). A judgment is here construed as something very general, of the form: ⟨a is F⟩. Given the generality of the account, it seems fairly clear that the ability to make—to determine the truth of—some such judgments is necessary for autonomous systems (even when this ability is not implemented in the form of a per se symbolic reasoner). Insofar as this is true—and given that Russell's principle is correct (I will not delve into Evans' interesting and convincing defense)—any autonomous system must know (or have the ability to discover) which thing in the world ⟨a⟩ is.

This hardly seems objectionable. The trouble, as Evans himself admits, is in spelling out what such knowledge amounts to. He suggests that the condition for knowing which thing ⟨a⟩ is might be met by an agent who: (1) possesses the knowledge of some discriminating feature of ⟨a⟩, or (2) has the ability to locate ⟨a⟩ in her vicinity, or (3) has the capacity to recognize ⟨a⟩, that is, the disposition to identify one (and only one) object as ⟨a⟩. Of course, even this specification of conditions leaves ample room for alternate interpretations (Evans spends some time on an effective critique of the photograph and causal theories of reference, demonstrating the inadequacy of their versions of the above criteria) but it does neatly and naturally suggest three varieties of reference deserving of further investigation: (1) information-based reference, (2) demonstrative reference, and (3) recognition-based reference, to which list Evans adds some other items, of which self-reference is the most important.

Taking each in turn, and roughly, an information-based thought about ⟨a⟩ "is the result of a belief about how the world is which the subject has because he has received information (or misinformation) from the object" (p. 121). In this case, the reference is to the object from which the information derives, even in the case where that information is mistaken, as in the famous case of referring to 'the man holding the champagne', whose glass is in fact full of sparkling cider. The paradigm case of demonstrative reference is the simple 'this', but also includes 'that', 'here', 'there', and all like descriptionless, indexical identifications. Finally (I shall here ignore self-reference, although Evans' account of it is interesting, and the relation he describes between 'here'-thoughts and 'I'-thoughts is central to his overall account) recognition-based reference deals with the case where an agent refers to an object previously encountered and remembered. Evans writes: "[I]f a subject is disposed to identify a particular object as the object of his thought, and in so doing is exercising a genuine recognitional capacity stemming from the encounter or encounters from which the memory information that saturates the thought derives, then, it seems to me, that object is the object of his thought, irrespective of whether or not it can be identified by means of any descriptions which the subject might otherwise have" (p. 269).

It is likely that Evans' discussions of demonstrative and recognition-based reference will have the most immediate relevance to those involved in understanding anchoring. And in this regard it is worth mentioning what I take to be Evans' greatest strength, considered from the standpoint of one interested in the behavior of autonomous, embodied agents: his insistence on situating reference in the larger context of being and acting in the world. I am impressed in particular with his argument that demonstrative reference requires of the agent awareness of an ego-centered space within which (and in terms of which) experience is instantiated and actions effected. Consider, in this regard, the difference between the judgments ⟨There's a fire here⟩ and ⟨There's a fire there⟩, or ⟨There's a dollar here⟩. Surely successful anchoring has not been displayed by a system that does not react differently in each case. That is, it is not enough to tag an object with an arbitrary symbol, and maintain this connection (although doing even this is not without its challenges!); one must connect with the right symbol, in the right way, so as to support appropriate reasoning about, and reaction to, the objects of the world.

In addition to recommending *The Varieties of Reference* as the single best philosophical resource for those interested in this immense project, I have also compiled a brief bibliography of core readings [1–12], and a longer list of other useful and important work [13–64]. It is my hope that the collective encounter with these works can help build a Lingua Franca of anchoring, without which the collaborative effort required to advance understanding in this difficult area will be much hindered.

Core readings

[1] T. Burge, Individualism and the Mental, Midwest Studies in Philosophy IV, University of Minnesota Press, Minneapolis, MN, 1979.

[2] K. Donellan, Reference and definite descriptions, in: A.P. Martinich (Ed.), The Philosophy of Language, Oxford University Press, Oxford, 1985.

[3] G. Evans, The Varieties of Reference, Oxford University Press, Oxford, 1982.

[4] G. Frege, On sense and meaning, in: A.P. Martinich (Ed.), The Philosophy of Language, Oxford University Press, Oxford, 1985.

[5] D. Kaplan, Demonstratives, in: A.P. Martinich (Ed.), The Philosophy of Language, Oxford University Press, Oxford, 1985.

[6] S. Kripke, Naming and Necessity, Harvard University Press, Cambridge, MA, 1980.

[7] A.P. Martinich (Ed.), The Philosophy of Language, Oxford University Press, Oxford, 1985.

[8] J. McDowell, Singular thought and the extent of inner space, in: P. Pettit, J. McDowell (Eds.), Subject, Thought, and Context, Oxford University Press, Oxford, 1986.

[9] J. McDowell, On the sense and reference of a proper name, Mind 86 (1977) 159–185.

[10] J. Perry, The Problem of the Essential Indexical, Oxford University Press, Oxford, 1993.

[11] H. Putnam, The meaning of meaning, in: Mind, Language and Reality, Cambridge University Press, Cambridge, 1975.

[12] P.F. Strawson, On referring, in: A.P. Martinich (Ed.), The Philosophy of Language, Oxford University Press, Oxford, 1985.

Extended bibliography

[13] F. Ackerman, Content, character, and nondescriptive meaning, in: J. Almog, J. Perry, H. Wettstein (Eds.), Themes from Kaplan, Oxford University Press, Oxford, 1989.

[14] F. Ackerman, Proper names, propositional attitudes, and nondescriptive connotations, Philosophical Studies 35 (1979) 55–69.

[15] L. Addis, Intrinsic reference and the new theory, in: P.A. French, T.E. Uehling, H.K. Wettstein (Eds.), Midwest Studies in Philosophy XIV: Contemporary Perspectives in the Philosophy of Language, University of Minnesota Press, Minneapolis, MN, 1990.

[16] K. Akins, Of sensory systems and the aboutness of mental states, Journal of Philosophy 93 (7) (1996) 337–372.

[17] K. Bach, Indexical content, in: Routledge Encyclopedia of Philosophy, Routledge, New York, 1998.

[18] K. Bach, Thought and Reference, Oxford University Press, Oxford, 1987.

[19] K. Bach, Thought and object: *de re* representations and relations, in: M. Brand, R.M. Harnish (Eds.), The Representation of Knowledge and Belief, University of Arizona Press, Tucson, AZ, 1986.

[20] T. Blackburn, The elusiveness of reference, in: P.A. French, T.E. Uehling, H.K. Wettstein (Eds.), Midwest Studies in Philosophy XII: Realism and Antirealism, University of Minnesota Press, Minneapolis, MN, 1988.

[21] D. Braun, Empty names, Noûs 32 (4) (1993).

[22] T. Burge, Reference and proper names, Journal of Philosophy 70 (1973) 425–439.

[23] H.N. Castañeda, Direct reference, the semantics of thinking, and guise theory, in: J. Almog, J. Perry, H. Wettstein (Eds.), Themes from Kaplan, Oxford University Press, Oxford, 1989.

[24] M. Devitt, Against direct reference, in: P.A. French, T.E. Uehling, H.K. Wettstein (Eds.), Midwest Studies in Philosophy XIV: Contemporary Perspectives in the Philosophy of Language, University of Minnesota Press, Minneapolis, MN, 1990.

[25] M. Devitt, Singular terms, Journal of Philosophy 71 (1974) 183–205.

[26] K. Donnellan, Proper names and identifying descriptions, in: S.P. Schwartz (Ed.), Naming, Necessity and Natural Kinds, Cornell University Press, Ithaca, NY, 1977.

[27] F. Dretske, The intentionality of cognitive states, in: W. Lycan (Ed.), Mind and Cognition, 2nd ed., Basil Blackwell, Oxford, 1999.

[28] F. Dretske, Aspects of cognitive representation, in: M. Brand, R. Harnish (Eds.), The Representation of Knowledge and Belief, University of Arizona Press, Tucson, AZ, 1986.

[29] M. Dummett, Frege: Philosophy of Language, 2nd ed., Harvard University Press, Cambridge, MA, 1981.

[30] M. Dummett, The Interpretation of Frege's Philosophy, Harvard University Press, Cambridge, MA, 1981.

[31] G. Evans, The causal theory of names, in: A.P. Martinich (Ed.), The Philosophy of Language, Oxford University Press, Oxford, 1985.

[32] G. Evans, Understanding Demonstratives, Collected Papers, Oxford University Press, Oxford, 1985.

[33] J. Fodor, A theory of content, in: W. Lycan (Ed.), Mind and Cognition, 2nd ed., Basil Blackwell, Oxford, 1999.

[34] G. Forbes, Cognitive architecture and the semantics of belief, in: P.A. French, T.E. Uehling, H.K. Wettstein (Eds.), Midwest Studies in Philosophy XIV: Contemporary Perspectives in the Philosophy of Language, University of Minnesota Press, Minneapolis, MN, 1990.

[35] I. Hacking, Representing and Intervening, Cambridge University Press, Cambridge, 1983.

[36] I. Hacking, Why Does Language Matter to Philosophy? Cambridge University Press, Cambridge, 1975.

[37] T. Horgan, J. Tienson, The intentionality of phenomenology and the phenomenology of intentionality, in: D. Chalmers (Ed.), Philosophy of Mind: Classical and Contemporary Readings, Oxford University Press, Oxford, 2002.

[38] S. Kripke, On rules and private language, in: A.P. Martinich (Ed.), The Philosophy of Language, Oxford University Press, Oxford, 1985.

[39] S. Kripke, Speaker's reference and semantic reference, in: A.P. Martinich (Ed.), The Philosophy of Language, Oxford University Press, Oxford, 1985.

[40] C. Li, Natural kinds: direct reference, realism and the impossibility of necessary, a posteriori truth, Review of Metaphysics 47 (1993) 261–276.

[41] K. Ludwig, Singular thought and the Cartesian theory of mind, Noûs 30 (4) (1996).

[42] J. McDowell, Mind and World, Harvard University Press, Cambridge, MA, 1994.

[43] C. McGinn, The mechanism of reference, Synthese 49 (1981) 157–186.

[44] M. McKinsey, Names and intentionality, Philosophical Review 87 (1978) 171–200.

[45] R. Millikan, Biosemantics, in: W. Lycan (Ed.), Mind and Cognition, 2nd ed., Basil Blackwell, Oxford, 1999.

[46] R. Millikan, White Queen Psychology and Other Essays for Alice, MIT Press, Cambridge, MA, 1995.

[47] R. Millikan, Language, Thought and Other Biological Categories, MIT Press, Cambridge, MA, 1984.

[48] M. O'Donovan-Anderson, Content and Comportment: On Embodiment and the Epistemic Availability of the World, Rowman and Littlefield, Lanham, 1997.

[49] D. Perlis, Consciousness as self function, Journal of Consciousness Studies 4 (5–6) (1997) 509–525.

[50] D. Perlis, Putting one's foot in one's head—Part II. How?, in: E. Dietrich (Ed.), From Thinking Machines to Virtual Persons: Essays on the Intentionality of Computers, Academic Press, New York, 1994.

[51] D. Perlis, Putting one's foot in one's head—Part I. Why? Noûs 25 (1991) 425–455.

[52] J. Perry, Indexicals and demonstratives, in: B. Hale, C. Wright (Eds.), A Companion to Philosophy of Language, Blackwell, Oxford, 1997.

[53] H. Putnam, Explanation and reference, in: Mind, Language, and Reality, Cambridge University Press, Cambridge, 1975.

[54] M. Richard, Indexicals, in: W. Bright (Ed.), International Encyclopedia of Linguistics, Oxford University Press, Oxford, 1992.

[55] F. Recanati, Direct Reference: From Language to Thought, Blackwell Press, Oxford, 1993.

[56] B. Russell, Descriptions, in: A.P. Martinich (Ed.), The Philosophy of Language, Oxford University Press, Oxford, 1985.

[57] S. Schiffer, Indexicals and the theory of reference, Synthese 49 (1981) 43–100.

[58] S. Schiffer, The basis of reference, Erkenntnis 13 (1978) 171–206.

[59] J. Searle, Proper names, in: A.P. Martinich (Ed.), The Philosophy of Language, Oxford University Press, Oxford, 1985.

[60] J. Searle, Intentionality, Cambridge University Press, Cambridge, 1983.

[61] R. Stalnaker, Indexical belief, Synthese 49 (1981) 129–151.

[62] P.F. Strawson, Meaning and truth, in: A.P. Martinich (Eds.), The Philosophy of Language, Oxford University Press, Oxford, 1985.

[63] P.F. Strawson, Individuals, Routledge, London, 1959.

[64] H. Wettstein, How to bridge the gap between meaning and reference, Synthese 84 (1984) 63–84.

Michael L. Anderson holds a Ph.D. in Philosophy from Yale University. He is currently a post-doctoral research associate at the University of Maryland, Institute for Advanced Computer Studies, where he studies the principles of self-correcting agency in real-world settings. His primary research goal is to show how reasoning and mental representation are grounded in physical structures and environmental interactions. Thus, for instance, he has argued that symbol anchoring requires the ability to interact with and manipulate the objects in one's environment. Current research includes an account of proprioception, detailing both how the particular structure of this aspect of our bodily experience is shaped by the physical and neural pathways which support it, as well as the role this bodily sense plays in structuring and supporting perception more generally.