

Externalism

Intermediate article

Robert A Wilson, University of Alberta, Edmonton, Canada

CONTENTS

What is externalism?

History of and arguments for externalism

Varieties of externalism

Externalism and cognitive science

Problems for externalism

Externalism is the view in the philosophy of mind that an individual's social or physical environment is partly determinative or constitutive of that individual's mental states.

WHAT IS EXTERNALISM?

The debate between individualists (or 'internalists') and anti-individualists (or 'externalists') is sometimes glossed in terms of whether psychological (or mental) states are 'in the head'. At first sight, that is likely to seem a trivial question: of course mental states are in the head ('Where else could they be?', as Robert Stalnaker (1989) asked). So we must first articulate a version of the issue that makes it clearer why externalism is a substantive, and thus potentially controversial, claim about the nature of the mind; we will then show why the (sometimes rarefied) debate over externalism has some important implications for how the mind should be studied with the cognitive sciences.

Consider the question of whether the character of an agent's environment plays some crucial role in determining or fixing the nature of that agent's mind. A natural thought (one, in fact, common to those who disagree about the answer to this question) would be that agents causally interact with their world, gathering information about it through their senses, and so the nature of their minds, in particular what their thoughts are about, is in part determined by the character of their world. Thus, the world is a causal determinant of one's thoughts, and hence one's mind. That is, the world is a contributing cause of the content, or intentionality, of one's mind, that is, what one perceives and thinks about. This is just to say that the content of one's mind is not causally isolated from one's environment. Individualists and externalists in the philosophy of mind agree on this much. What separates them is the question of whether there is some deeper sense in which the nature of the mind is

determined by the character of the individual's world. It is this 'deeper sense' of world-mind determination that we need to articulate.

We can approach this task by extending the brief discussion above of the idea that the content of the mind is in part causally determined by the agent's environment to explore the conditions under which a difference in the world implies a difference in the mind. Individualists hold that this is so just in case that difference in the world makes some corresponding change to what occurs inside the boundary of the individual; externalists deny this, thereby allowing for the possibility that individuals who are identical with respect to all of their intrinsic features could none the less have psychological or mental states with different contents. And, assuming that mental states with different contents are *ipso facto* different types or kinds of states, this implies that an individual's intrinsic properties do not determine or fix that individual's mental states.

This provides us with another, more precise way of articulating the difference between externalism and individualism about the mind. Individualists claim, and externalists deny, that what occurs inside the boundary of an individual metaphysically determines the nature of that individual's mental states. The individualistic determination thesis, unlike the causal determination thesis, expresses a view about the nature or essence of mental states, and identifies a way in which, despite their causal determination by states of the world, mental states are autonomous or independent of the character of the world beyond the individual. What individualism implies is that two individuals who are identical in all their intrinsic respects must have the same psychological states. It is the modal aspect to this implication that makes supervenience an appropriate concept to use in stating individualism more precisely: an individual's psychological states must supervene on the intrinsic physical states of that individual.

This implication, and indeed the debate over externalism, is often made more vivid through the fantasy of doppelgangers, pairs of individuals identical at the molecular level, and the corresponding fantasy of 'Twin Earth'. We will return to these fantasies later in this article.

HISTORY OF AND ARGUMENTS FOR EXTERNALISM

Hilary Putnam's paper "The meaning of "meaning"" (1975) introduced both of the above fantasies in the context of a discussion of the meaning of natural language terms. Putnam was concerned to show that 'meaning' does not and cannot jointly satisfy two theses that it was often taken to satisfy by then-prevalent theories of natural language reference: the thesis that the meaning of a term is what determines its reference; and the thesis that meanings are 'in the head', where this phrase should be understood as making a claim of the type identified above about the metaphysical determination of meanings. These theses typified descriptive theories of reference – first formulated by Frege and Russell – according to which the reference of a term is determined by the descriptions that a speaker attaches to that term. In attacking this view and its presuppositions Putnam focused on terms denoting natural kinds, such as 'water' and 'tiger', but he intended his attack, and his subsequent alternative theory of natural language reference, the causal theory of reference, to be more general than this.

Consider an ordinary individual, Oscar, who lives on Earth and interacts with water in the ways that most of us do: he drinks it, washes with it, and sees it falling from the sky as rain. Oscar, who has no special chemical knowledge about the nature of water, will associate a range of descriptions with his term 'water': it is a liquid that one can drink, that is used to wash, and that falls from the sky as rain. In a descriptive theory of reference, these descriptions determine the reference of Oscar's term 'water'. That is, Oscar's term 'water' refers to whatever it is in the world that satisfies the set of descriptions he attaches to the term. And since those descriptions are 'in the head', natural language reference on this view is individualistic. But now, to continue Putnam's argument, imagine a molecule-for-molecule doppelganger of Oscar, Oscar*, who lives on a planet just like Earth in all respects but one: the substance that people drink, wash with, and see falling from the sky is not water (i.e., H₂O), but a substance with a

different chemical structure, XYZ. Call this planet 'Twin Earth'. This substance, XYZ, is called 'water' on Twin Earth, and Oscar*, as a doppelganger or twin of Oscar, has the same beliefs about it as Oscar has about water on Earth. (Recall that Oscar, and thus Oscar* has no special knowledge of the chemical structure of water.) Twin Earth has what we might call 'twin water' or 'twater' on it, not water, and it is twater that Oscar* interacts with, not water – after all, there is no water on Twin Earth. Given that Oscar's term 'water' refers to or is about water, then Oscar*'s term 'water' refers to or is about twater. That is, they have natural language terms that differ in their meaning, assuming that reference is at least one aspect of meaning. But, by hypothesis, Oscar and Oscar* are doppelgangers, and so are identical in all their intrinsic properties, and so are identical with respect to what is 'in the head'. Thus, Putnam argues, the meanings of the natural language terms that Oscar and Oscar* use are not metaphysically determined by what is in the head of the individual using those terms.

Putnam was attacking a tradition of thinking about language which was, in terms that Putnam appropriated from Carnap's *Aufbau*, 'methodologically solipsistic': it treated the meanings of natural language terms, and language more generally, in ways that supposed that the world beyond the individual language user did not exist. Since Putnam's chief point is one about natural language terms and the relationship of their semantics to what is inside the head, one needs at least to extend his reasoning from language to thought to arrive at a position that denies individualism about the mind itself. But given the tradition to which he was opposed, such an extension might be thought to be relatively easy, since in effect those in the tradition of methodological solipsism – from Descartes to Brentano, to Russell, to Husserl, to Carnap – conceived of natural languages and their use in psychological terms.

The term 'individualism' itself, as well as a series of thought experiments which made a case for externalism and in some ways paralleled Putnam's 'Twin Earth' thought experiment, was introduced by Tyler Burge in his paper 'Individualism and the Mental' (1979). Burge identified individualism as an overall conception of the mind prevalent in modern philosophical thinking at least since Descartes, and argued that our commonsense psychological framework for explaining behavior, our 'folk psychology', was externalist. Burge was explicit in making a case against individualism that did not turn on any controversial claims about the

semantics of natural kind terms: he developed his argument using agents with thoughts about arthritis, sofas, and contracts, so did not presuppose any type of scientific essentialism about natural kinds. Like Putnam's argument, however, Burge's argument does presuppose some views about natural language understanding. The most important of these is that we can and do have incomplete understanding of many of the things that we have thoughts about and for which we have natural language terms.

VARIETIES OF EXTERNALISM

Putnam's externalism is sometimes characterized as a form of 'physical' externalism, in contrast to Burge's 'social' externalism. According to Putnam, the character of the physical world (e.g. the nature of water itself) in part metaphysically determines the content of one's mind; while according to Burge the character of the social world (e.g. the nature of one's linguistic community) does so. While these terms may serve as a useful reminder of one way in which these two views differ, we should also keep in mind the 'social' aspect of Putnam's view of natural language: his linguistic division of labor. Important to both views is the idea that language users and psychological beings depend and rely on one another in ways that are reflected in our everyday, common-sense ways of thinking about language and thought. Thus, on both views, there is a social aspect to the nature of meaning and thought. This is partly what justifies the label *anti-individualism* for both of them.

There is another division between varieties of externalism that turns on their relationship with the cognitive sciences. Taking their cue from the linguistic emphasis of the Putnam-Burge arguments for externalism, some externalists, such as McDowell (1986) and Pettit (1983), have developed their views within a broadly Wittgensteinian tradition of thinking about the mind, one that views externalism as latent in our ordinary language and the 'language games' that we play with mental terms. But other philosophers, such as Segal (1989), Egan (1992, 1995) and Shapiro (1993, 1997), locate the debate between individualists and externalists within the areas of the philosophy of science that concern cognition. These philosophers have probed general notions, such as that of computation and mental representation, central to understanding cognition, as well as exploring specific research programmes within cognitive science for their relationship to the externalism debate.

EXTERNALISM AND COGNITIVE SCIENCE

At around the time when individualism was coming under attack from the externalism of Putnam and Burge, it was also being defended (e.g., by Fodor (1980) and Stich (1983)) as a view of the mind particularly apt for a genuinely scientific approach to understanding the mind, especially of the type that was being articulated within the nascent interdisciplinary field of cognitive science. For these defenders of individualism, there was something suspiciously unnaturalistic about the Putnam-Burge arguments, as well as something about their conclusions that seemed anti-scientific. Part of the defence of individualism and the corresponding attack on externalism turned on what we may call the 'cognitive science gesture': the claim that, as contemporary empirical work on cognition indicated, any truly scientific understanding of the mind would need to be individualistic, and thus could not be externalist.

Common to both Fodor's and Stich's views of cognitive science is the idea that an individual's psychological states should be *bracketed off* from the environments beyond the head in which individuals find themselves. Unlike Putnam and Burge in the papers discussed above, Fodor and Stich have focused on the relevance of individualism to psychological explanation, and have used their respective principles to argue for substantive conclusions about the scope and methodology of psychology and the cognitive sciences. Fodor (1980) contrasted a 'methodologically solipsistic' psychology with what he called a naturalistic psychology, and argued that since the latter (in which he included J. J. Gibson's approach to perception, learning theory, and the naturalism of William James) was unlikely to lead to reliable research in psychology, methodological solipsism was the only fruitful psychological framework for understanding cognition (see also Fodor, 1987). One implication of Fodor's position is that the relevant notion of content for cognitive science is *narrow content* (White, 1991). Stich (1978) argued for a syntactic or computational theory of mind that made no essential use of the notion of intentionality or mental content. Thus he deployed individualism in defence of an eliminativist view of content (see also Stich, 1983).

Although the 'cognitive science gesture' is a gesture rather than a solid argument that appeals to empirical data, it is not an empty gesture. Although Fodor's and Stich's arguments have not won widespread acceptance either by philosophers or by

cognitive scientists, they have struck a chord with the latter, perhaps not surprisingly since the dominant research traditions in cognitive science have been at least implicitly individualistic. Relatively explicit commitments to an individualistic view of aspects of cognitive science include Chomsky's (1986, 1995) use of the distinction between two conceptions of language (the 'I'-language and the 'E'-language, for 'internal' and 'external' respectively), Jackendoff's (1991) related, general distinction between 'psychological' and 'philosophical' conceptions of the mind, and Cosmides and Tooby's (1994) more recent emphasis on the constructive nature of our internal, evolutionary-specialized cognitive modules.

Two further points in brief. First, there is a large literature that discusses Marr's (1982) theory of vision in connection with externalism, beginning with Burge (1986), which raises and discusses many issues relevant to assessing the relevance of externalism for cognitive science (see Wilson, 2002). Second, an important way to develop the externalist view of the mind is to pursue its connection to 'embodied' or 'embedded' approaches to cognition (e.g. Clark, 1997; Hutchins, 1995; McClamrock, 1995). These approaches question business as usual in cognitive science, and their development is likely to be fruitful for the study of cognition.

PROBLEMS FOR EXTERNALISM

Two important problems for externalism, especially for practicing cognitive scientists, are its perceived incompatibility with the insights of the computational and representational theories of mind. The former of these theories holds that mental processing is a form of computation. The latter theory holds that we interact with the world perceptually and behaviorally through internal mental representations of how the world is (as the effects of perceiving) or how the world should be (as instructions to act). Provided that the appropriate internal representational states of the organism remain fixed, the organism's more peripheral causal involvement with its environment is irrelevant to cognition, since that involvement cannot alter the internal mental states that represent that environment. Wide computationalism (Wilson, 1994 and 1995) is one response to the first of these problems; Burge's (1986) interpretation of Marr's theory of vision is one response to the second. (For replies, see Segal, 1989 and 1998, and Matthews, 1988.)

Externalism has also been thought to give rise to various related, but more purely philosophical,

problems; for example, by failing to make sense of the notion of mental causation (e.g. Block, 1986), or misconstruing the role of causal powers in psychological taxonomy (Fodor, 1987). Connecting such objections of individualism with the methodological formulations that have influenced cognitive science, it has been claimed that individualism provides a minimal constraint needed to arrive at psychological explanations that locate the mind suitably in the physical world, a psychology that taxonomizes its entities by their causal powers.

A further philosophical problem for externalism concerns its compatibility with a cluster of related epistemological issues: those concerning self-knowledge, a-priori knowledge, and scepticism.

Basic to self-knowledge is knowledge of one's own mind, and traditionally this knowledge has been thought to involve some form of privileged access to one's own mental states. This notion of epistemic privilege has been developed in a number of ways, all of which share the idea that there is an asymmetry between knowledge of one's own mind and knowledge of the minds of others and of other things in the world. (Indicative of the depth of these asymmetries in modern philosophy is the fact that an introduction to epistemology that reflects on scepticism will likely introduce the *problem of other minds* and the *problem of our knowledge of the external world*, but not the corresponding *problem of self-knowledge*.) Scepticism about one's own mind has seemed to be precluded by the very nature of self-knowledge.

Individualistic conceptions of the mind have seemed well suited to making sense of first-person privileged access and the asymmetry between self-knowledge and knowledge of the mental states of others. If mental states are individuated in abstraction from the environment 'beyond the individual', then there seems to be no problem in understanding how the process of introspection – turning our mind's eye inwards, to use a common metaphor – reveals the content of those states. To invoke the Cartesian fantasy in a way that brings out the asymmetry between self-knowledge and other forms of knowledge: even if there were an evil demon who deceived me about the existence of an external world – including the existence of other people with mental states like mine – the one thing that I could be sure about would be that I am having experiences with a certain content. As it is sometimes put, even if I could be deceived about whether there is really a tree in front of me and thus about whether I am actually seeing a tree, I can't be deceived about whether it seems to me that I am

seeing a tree. Thus individualism seems to confer a certain epistemic security on first-person knowledge of one's own mental states which the corresponding third-person knowledge lacks.

Externalism, by contrast, poses a *prima facie* problem for even the more modest forms of first-person privileged access, and has even been thought to call into question the possibility of any form of self-knowledge. For externalism claims that what mental states are is metaphysically determined, in part, by the nature of the world beyond the boundary of the subject of those states. Thus it would seem that in order to know what one is thinking (that is, to know the content of one's mental states), one would have to know something about the world beyond one's self. But this would be to assimilate our first-person knowledge of our own minds to our knowledge of other things, and so to deny any privileged access that self-knowledge might have. It implies that in order to know my own mind I need to know facts (perhaps difficult to discern) about the nature of the physical or social world in which I live; and so also suggests that in a range of ordinary cases where we might unreflectively attribute self-knowledge, I don't actually have self-knowledge.

We can express the problem in another way, which abstracts away from the differences between specific accounts of privileged access and specific accounts of externalism. Knowledge of one's own mental states, whether it be infallible, incorrigible, self-intimating, introspective, or a priori, has a special character. Knowing one's own mental states involves, *inter alia*, knowing their contents. Now, according to externalism, the contents of a subject's mental states are metaphysically determined, in part, by facts about his or her physical or social environment. Knowledge of these facts, however, does not have this special character. How is the special character of self-knowledge compatible with the non-special character of worldly knowledge, given the dependence of the former on the latter?

The problem can be represented as a supposedly inconsistent triad of propositions (see also McKinsey, 1991). Let *P* stand for the contents of our mental states; let *E* stand for facts about the environment; and let 'by introspection' stand for the special character of self-knowledge.

We know *P* by introspection.
(Self-Knowledge) (1)

P are metaphysically determined in part
by *E*. (Externalism) (2)

E are not known by introspection.
(Common Sense) (3)

The claim is that one of these three propositions must be rejected. If we reject Self-Knowledge, then we give up on the idea that we have privileged access to our own minds; if we reject Externalism, then we return to individualism; and if we reject Common Sense, then we make a strange and implausible claim about our knowledge of the physical or social world.

References

- Block N (1986) Advertisement for a semantics for psychology. In: French P, Uehling T and Wettstein H (eds) *Midwest Studies in Philosophy*, vol. X 'Philosophy of Mind'. Minneapolis, MN: University of Minnesota Press.
- Burge T (1979) Individualism and the mental. In: French P, Uehling T and Wettstein H (eds) *Midwest Studies in Philosophy*, vol. IV 'Metaphysics'. Minneapolis, MN: University of Minnesota Press.
- Burge T (1986) Individualism and psychology. *Philosophical Review* 3–45.
- Chomsky N (1986) *Knowledge of Language*. New York, NY: Praeger.
- Chomsky N (1995) Language and nature. *Mind* 1–61.
- Clark A (1997) *Being There: Putting Mind, Body and World Together*. Cambridge, MA: MIT Press.
- Cosmides L and Tooby J (1994) Foreword to Baron-Cohen S (1994) *Mindblindness*. Cambridge, MA: MIT Press.
- Egan F (1992) Individualism, computation, and perceptual content. *Mind* 443–459.
- Egan F (1995) Computation and content. *Philosophical Review* 181–203.
- Fodor JA (1980) Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Science* 63–73. [Reprinted in: Fodor JA (1981) *Representations*. Sussex, UK: Harvester Press.]
- Fodor JA (1987) *Psychosemantics*. Cambridge, MA: MIT Press.
- Hutchins E (1995) *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Jackendoff R (1991) The problem of reality. [Reprinted in: Jackendoff R (1992) *Languages of the Mind*. Cambridge, MA: MIT Press.]
- Marr D (1982) *Vision: A Computational Approach*. San Francisco, CA: Freeman.
- Matthews R (1988) Comments on Burge. In: Grimm and Merrill (eds) *Contents of Thought*, pp 77–86.
- McClamrock R (1995) *Existential Cognition: Computational Minds in the World*. Chicago, IL: University of Chicago Press.
- McDowell J (1986) Singular thought and the extent of inner space. In: Pettit P and McDowell J (eds) *Subject, Thought, and Context*. Oxford: Oxford University Press.

- McKinsey M (1991) Anti-individualism and privileged access. *Analysis* 51: 9–16.
- Pettit P (1983) Wittgenstein, individualism and the mental. In: *Epistemology and the Philosophy of Science: Proceedings of the Seventh International Symposium*. Vienna: Holder-Pichler-Tempsky.
- Putnam H (1975) The meaning of 'meaning'. In: Gunderson K (ed.) *Language, Mind, and Knowledge. Minnesota Studies in the Philosophy of Science*, vol. 7. Minneapolis, MN: University of Minnesota Press. [Reprinted in: Putnam H (1979) *Mind, Language, and Reality*. New York, NY: Cambridge University Press.]
- Segal G (1989) Seeing what is not there. *Philosophical Review* 98: 189–214.
- Segal G (1997) Review of R. A. Wilson, Cartesian Psychology and Physical Minds. *British Journal for the Philosophy of Science* 48: 151–157.
- Shapiro L (1993) Content, kinds, and individualism in Marr's theory of vision. *Philosophical Review* 102: 489–513.
- Shapiro L (1997) A clearer vision. *Philosophy of Science* 64: 131–153.
- Stalnaker RC (1989) On what's in the head. In: Tomberlin J (ed.) *Philosophical Perspectives*, vol. III. Atascadero, CA: Ridgeview.
- Stich S (1978) Autonomous psychology and the belief-desire thesis. *Monist* 61: 573–591.
- Stich S (1983) *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.
- White S (1991) *The Unity of the Self*. Cambridge, MA: MIT Press.
- Wilson RA (1994) Wide computationalism. *Mind* 103: 351–372.
- Wilson RA (1995) *Cartesian Psychology and Physical Minds: Individualism and the Sciences of the Mind*. New York, NY: Cambridge University Press.
- Wilson RA (2002) Individualism. In: Warfield T and Stich S (eds) *The Blackwell Guide to the Mind*. Oxford: Blackwell Publishers.
- Burge T (1988a) Cartesian error and the objectivity of perception. In: Grimm and Merrill (eds) *Contents of Thought*, pp. 62–76. Tucson, AZ: University of Arizona Press. Also in: Pettit P and McDowell J (eds) (1986) *Subject, Thought, and Context*, pp. 117–136. Oxford, UK: Oxford University Press.
- Burge T (1988b) Individualism and self-knowledge. *Journal of Philosophy* 85: 649–663.
- Burge T (1989) Individuation and causation in psychology. *Pacific Philosophical Quarterly* 70: 303–322.
- Davies M (1991) Individualism and perceptual content. *Mind* 100: 461–484.
- Devitt M (1990) A narrow representational theory of mind. In: Lycan W (ed.) *Mind and Cognition: A Reader*. New York, NY: Blackwell.
- Egan F (1999) In defense of narrow mindedness. *Mind and Language* 14: 177–194.
- Fodor JA (1982) Cognitive science and the Twin Earth problem. *Notre Dame Journal of Formal Logic* 23: 98–118.
- van Gulick R (1989) Metaphysical arguments for internalism and why they don't work. In: Silvers S (ed.) *Representation*. Dordrecht, The Netherlands: Kluwer.
- Ludlow P and Madin N (eds) (1998) *Externalism and Self-Knowledge*. Palo Alto, CA: CSLI Publishers.
- McCulloch G (1995) *The Mind and Its World*. New York, NY: Routledge.
- McGinn C (1989) *Mental Content*. New York, NY: Blackwell.
- Millikan R (1993) *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.
- Patterson S (1991) Individualism and semantic development. *Philosophy of Science* 58: 15–35.
- Rowlands M (1999) *The Body in Mind: Understanding Cognitive Processes*. New York, NY: Cambridge University Press.
- Segal G (2000) *A Slim Book about Narrow Content*. Cambridge, MA: MIT Press.
- Walsh DM (1999) Alternative individualism. *Philosophy of Science* 66: 628–648.
- Wilson RA (2000a) The mind beyond itself. In: Sperber D (ed.) *Metarepresentation*. New York, NY: Oxford University Press.
- Wilson RA (2000b) Some problems for 'alternative individualism'. *Philosophy of Science* 67: 671–679.
- Wilson RA (2001) Two views of realization. *Philosophical Studies* 104: 1–30.
- Woodfield A (ed.) (1982) *Thought and Object: Essays on Intentionality*. Oxford, UK: Oxford University Press.

Further Reading

- Adams F, Drebusenko D, Fuller G and Stecker R (1990) Narrow content: Fodor's folly. *Mind and Language* 5: 213–229.
- Burge T (1982a) Other bodies. In: Woodfield, 1982, pp. 97–120.
- Burge T (1982b) Two thought experiments reviewed. *Notre Dame Journal of Formal Logic* 23: 284–293.